OXFORD

Elements of Sonata Theory

Norms, Types, and Deformations in the Late-Eighteenth-Century Sonata

James Hepokoski
Warren Darcy

ELEMENTS *of* SONATA THEORY

This page intentionally left blank

ELEMENTS *of* SONATA THEORY

Norms, Types, and Deformations in the Late-Eighteenth-Century Sonata

James Hepokoski
Warren Darcy

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further
Oxford University's objective of excellence
in research, scholarship, and education.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

First published as an Oxford University Press paperback, 2011

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016

www.oup.com

Library of Congress Cataloging-in-Publication Data
Hepokoski, James A. (James Arnold), 1946–
Elements of sonata theory: norms, types, and deformations in the
late-eighteenth-century sonata / James Hepokoski, Warren Darcy.
p. cm.
Includes bibliographical references and index
ISBN-13: 978-0-19-977391-6
1. Sonata form. 2. Instrumental music—18th century—Analysis, appreciation.
I. Darcy, Warren. II. Title
MT62.H46 2006
784.18'3'09033—dc22 2005006674

1 3 5 7 9 8 6 4 2

Printed in the United States of America
on acid-free paper

Preface

This book offers a fresh approach to one of the most familiar topics in the field of music: the study of sonata-form movements and the larger workings of multimovement sonatas, symphonies, and chamber music of the "early classical" and "classical" period. While remaining in dialogue with the several current approaches to this subject, it provides something different, and from time to time it challenges established views of the sonata. Both of the authors have been leading classes and seminars in this method over the past decade at Yale University, Oberlin College Conservatory, and the University of Minnesota. Large portions of *Elements of Sonata Theory*, both in earlier incarnations and in this much-expanded one, have been required reading in these courses.

From one perspective the *Elements* is a research report, the product of our analyses of hundreds of individual movements by Haydn, Mozart, Beethoven, and many surrounding composers of the time (as well as later composers). In our work we have been looking for patterns within sonata-composition, for shared gestures, for ranges of options, for a sense of the typical. Our intention was to devise an explanation of how varying degrees of the normative can be altered, stretched, or occasionally overridden altogether to produce an individualized "deformation." To be sure, the theoretical discussions of eighteenth- and early- nineteenth-century writers are relevant (Koch, Galeazzi, Reicha, and others), and these insights are integrated into the book. Our preference, though, was to let the composers themselves teach us how sonatas work. Our method of understanding sonatas ("Sonata Theory") strikes a balance between inductively inferred norms and the unpredictability that one finds in these pieces.

Late-eighteenth-century sonatas are most productively heard within the context of a broad, flexible background-knowledge of what had come to be more or less standard compositional options at each point in the sonata. Any individual work within a genre (such as sonata form) interacts with the listener's (or composer's) expectations. Our book provides a detailed map of those expectations at that point in history. Not surprisingly, this turns out to be a complex matter. How can we know whether Haydn's choice here or Mozart's there was to be heard as normative, as strikingly unusual, or as something in between? And how can a composer's early choices influence the range of continuation-options down the road? Understanding any compositionally selected gesture requires an awareness of the backdrop of typical choices against which it was written and within whose world of norms the piece was to be grasped in the first place. The *Elements* seeks to fill in many details of that backdrop. This perspective has the advantage of permitting one to pass beyond the confines of the acoustic surface alone (what one literally hears, what is actually notated) in order to notice, for instance, which normative things might be absent. It could be that such absences—generically expected events that the composer might keep from happening within an individual work—should be understood as essential constituents of the piece's meaning.

This book divides into two large parts. In the first of these, chapters 1–15, we lay out the basics of the essential system, working section by section, zone by zone, through the most often encountered type of sonata form ("Type 3," with an exposition, development, and recapitulation), and considering also the differing implications within minor-mode sonata form and the multimovement sonata as a whole. In several of our analytical seminars earlier versions of chapters 1–15 alone served as the text. The discussions found in the second part, chapters 16–22, are more complex and extended, especially from chapter 17 onward. These chapters provide elaborately detailed studies of the other sonata formats of the period (Types 1, 2, 4, and 5). The increased intensity of these chapters is no accident. Confronting these differing formats at all—the "sonata without development" (Type 1), the "binary" sonata (Type 2, without a full recapitulation), the sonata-rondo (Type 4), and concerto first-movement form (Type 5)—throws one directly into the midst of ongoing debates and passionately held, sharply diverging views. Given the existing state of the discussion, we were obliged to present these thorny issues with an enhanced rigor, constructing step-by-step solutions to these often misconstrued matters and providing evidence and justification for our decisions along the way. Nowhere is this situation more evident than in the case of the first movements of Mozart's concertos. This Type 5 structure is the most difficult of the sonata types, and it is a topic concerning which even the most rudimentary features of terminology and sonata-form perception

have been ardently contested over the past century. Covering this problem adequately required four extended chapters (19–22)—virtually a separate monograph on the Mozartian Type 5 sonata, though one that is entirely dependent on one's grasp of the book's first eighteen chapters.

In addition to furnishing a new mode of analysis for the late-eighteenth-century instrumental repertory, the *Elements* also provides a foundation for considering works from the decades to come—late Beethoven, Schubert, Weber, Mendelssohn, Schumann, Liszt, Brahms, Bruckner, Strauss, Mahler, the "nationalist composers," and so on. As we point out from time to time, most of these sonata norms remained in place as regulative ideas throughout the nineteenth century, even as the whole sonata-form genre, with its various options, was continuously updated, altered, and further personalized with unforeseen accretions, startling innovations, and more radical deformations. (The "three-key expositions" sometimes found in Schubert and Brahms, for example—though surfacing in some earlier composers as well—seem to have been encouraged by the eighteenth-century expositional strategy of the "trimodular block" and its "apparent double medial caesuras." Similarly, the "de-energizing transition" and occasional suppression of the medial caesura in, say, Schumann or Brahms, surely emerged from the precedents of the "blocked medial caesura" coupled with "expanded caesura-fill" in Haydn and Mozart.)

What follows is a blend of musicological and music-theoretical thinking. What at first may seem to be a work of music theory turns out in the end to be a set of reflections on what sonata form is and how it can be understood to mean anything at all. In its drive to get to the bottom of things, Sonata Theory is informed by a not-always-tacit dialogue with current philosophy and literary criticism. While the book does not flaunt its intersections with certain strands of thought of the past decades—genre theory, phenomenology-oriented hermeneutics, reader (listener)-generated artistic texts, the slippage and dispersion of meanings once supposed to be unitary, and so on—the importance to us of those modes of thinking should be evident to most readers. (The more generalized axioms grounding our conceptual system and modes of inquiry are laid out in appendix 1 at the end of the book.) There are no tacit social agendas to our research of which we are aware, except that of seeking to understand what sonatas are and how they work. Still, Sonata Theory does have an interpretational, self-reflective, or philosophical tilt to it, an urge to explore a more fundamental questioning of this music's methods and purposes. We are committed to understanding musical practice not only as a self-contained technical language but also as a metaphor for human action or communication. We hope that our work will illuminate other perspectives and will open the investigation of sonata form and its diversified

meanings to questions of serious concern to a new, younger generation of musicians and scholars.

The musical examples in this book were created by Marcus Lofthouse, a recent graduate of the Oberlin Conservatory of Music, using Sibelius 3. Passages for solo piano are reproduced in full; string quartet excerpts are presented on two staves, but nothing has been omitted. Most orchestral passages are presented in two-stave reductions that eliminate some octave doublings but retain the melody and bass lines, accompaniment, and any inner voices or counterpoint, all in their original registers. Space limitations necessitated reducing most concerto passages to two staves, one for the soloist (usually piano), the other for the orchestral accompaniment. The solo passages are presented as completely as possible, eliminating only a few low doublings as well as those measures where the piano functions as a thoroughbass instrument. The orchestral passages required a bit more in the way of compromise, but the one-staff reductions do show all the essentials of the textures. The examples were checked against the most authoritative editions available. Although fidelity to the score is balanced by practical considerations of legibility, our aim has been to make these examples as complete, as faithful to the original, and as helpful as possible. The figures and tables in the various chapters were reproduced by Zachariah Victor, using Adobe Illustrator 10.

In referring to individual works in the text we normally use the full versions of the most widely known, easily recognizable titles and numberings (and even nicknames), even when those designations might be more popular—or customary—than scholarly. (Additionally, when we do not explicitly flag a key as "minor," we mean that it is major: "in D" means "in D major.") As all scholars of the period are aware, Mozart's "Symphony No. 39 in E-flat, K. 543" is not at all his thirty-ninth composed symphony—nor did the composer think of it in that way—although for a very long time it has been commonly referred to as that in standard discussions and shows no sign of even beginning to shake off this now "historically fixed" number. And merely to refer to the work, *de haut en bas*, as the Symphony in E-flat [Major] or only as K. 543, without any other identifying reference, could either oblige some readers to scurry off to Köchel-number lists or discourage them from trying to remember which piece this actually is. These issues are particularly noticeable in references to Mozart's piano concertos, which in the literature are often referred to only by Köchel number, sometimes accompanied by the key ("K. 488" alone or "Piano Concerto in A, K. 488," as opposed to our preferred—though not literally correct—"Piano Concerto No. 23 in A, K. 488"). We recognize the historical inaccuracies embedded in "No. 39" and "No. 23," but our intention, because we cite so many references to so many individual works, has been to simplify things for the reader. For similar reasons, within the text

proper our references to Köchel numbers are only to the familiar, "traditional" numbers, thereby avoiding the clutter and pedantic flavor of the double-descriptions that append the revised K. numbers as well, when such numbers exist. Thus instead of the scholarly precise "Piano Concerto in C, K. 415/387*b*" we prefer the more reader-friendly (albeit "incorrect" or not fully up-to-date with regard to the catalogue) "Piano Concerto No. 13 in C, K. 415." (No reader could possibly be confused by the absence of the much less familiar "updated number.") Not all of the K. numbers have these issues associated with them, but when they do, the dual number is provided in the index. Related issues and choices were made in citing the works of Haydn, Beethoven, Schubert, and others.

Any book of this scope is inevitably indebted to the many colleagues and students—too numerous to mention individually—with whom, along the way, we have shared information, proposed new ideas, developed concepts, and worked through analyses. We are grateful for all of these conversations and critiques, which have helped to shape our own thinking over the years. Apart from its use in academic classes and seminars, much of the first half of this book was "officially launched" at a workshop of the Mannes Institute for Advanced Studies in Music Theory (Institute on Musical Form) on June 24–27, 2004. This workshop was ably led by Daniel Harrison, and we thank him for doing so. We are also pleased to acknowledge the assistance and encouragement of Oxford University Press from the book's initial inception back in the mid- and late-1990s onward. We are grateful to Maribeth Payne, then music editor at Oxford, and to the group of anonymous reviewers that read and commented on an early version of this text around seven years ago. And we are indebted to the team at Oxford, Kimberley Robinson, Eve Bachrach, Robert Milks, Norman Hirschy, and others, who have been crucial in guiding this book through the production process and into print. Still additional thanks are due to Thomas Hepokoski, who helped to sustain this project to the end in important ways. Finally, we thank our immediate families—and especially our wives, Barbara and Marsha—for having the willingness, love, and patience to persevere through the seemingly endless sessions of our research, writing, and revisions. There may be, finally, light at the end of the tunnel.

This page intentionally left blank

Contents

This page intentionally left blank

Terms and Abbreviations

PAC = perfect authentic cadence (a phrase-concluding formula featuring V-I root-position bass motion; the upper voice ends on scale-degree $\hat{1}$ above the tonic chord)

IAC = imperfect authentic cadence (similar to PAC, but the upper voice ends on scale-degree $\hat{3}$ or $\hat{5}$ above the tonic chord)

HC = half-cadence (a cadence ending on an active V chord; this dominant chord will also end a phrase)

DC = deceptive cadence (V-vi, or V followed by any non-tonic chord containing $\hat{1}$, where an authentic cadence is expected)

V_T = a V that is tonicized; the dominant sounded as a key (as in second themes of major-mode expositions)

V_A = a V that is an active chord, not a key; the A stands for "active," and it indicates that the dominant is being sounded but not tonicized; instead, it implies a resolution to the existing or implied tonic.

C = closing zone (within an exposition, musical material following the EEC. Its internal modules are designated as C^1, C^2, etc.; in this case the superscript integers should be advanced only after a PAC.)

CF = caesura-fill (connective material, of variable length, bridging a caesura—either a medial caesura or a final caesura—to the next thematic module)

$C^{pre\text{-}EEC}$ = A "C"-like theme that occurs before the EEC proper within a continuous exposition. (Within a two-part exposition, such a theme is designated as S^C.)

CRI = coda-rhetoric interpolation (coda-like material interpolated shortly before the close of the recapitulatory rotation, which then resumes to complete the recapitulation proper)

DE = display episode (in a Type 5 sonata [concerto movement]

the solo-virtuosic closing portion, ending with an emphatic trill cadence, of S1 and S3—the solo exposition and solo recapitulation. The location of the display episode is usually included in the label, as in S1:\DE. See S1:\.)

EEC = essential expositional closure (within an exposition, usually the first satisfactory PAC that occurs within S and that proceeds onward to differing material. An immediate repetition of the melody or cadence—or certain other procedures, outlined in chapter 8—can defer this point to the next PAC.)

ESC = essential structural closure (within a recapitulation, usually the first satisfactory PAC that occurs within S and that proceeds onward to differing material. Like the EEC, the ESC can also be deferred through certain procedures to the next PAC. The ESC is normally the recapitulation's parallel point to the exposition's EEC, although exceptions do exist.)

FS = *Fortspinnung* modules (usually in the continuous-exposition context of TR⇒FS)

MC = medial caesura (within an exposition, I:HC MC represents a medial caesura built around the dominant of the original tonic; V:HC MC represents an MC built around V/V; etc. The presence of an MC identifies the exposition-type as two-part—the most common type—and leads directly to an S theme. In nearly all cases, if there is no MC, there is no S. Cf. the alternative, TR⇒FS.)

MMS = multimodular S (an S that tracks through two or more different, often contrasting ideas—$S^{1.1}$, $S^{1.2}$, and so on—before driving to its first satisfactory PAC with a cadential module. The numbers after the decimal point—the "decimal designators"—provide a method of labeling and identifying these separate modules. A *trimodular S* is particularly common: see TMS.)

P = primary-theme zone (whose individual modules may be described as $P^{1.1}$, $P^{1.2}$, etc. A module that precedes or sets up what is taken to be the "P-theme proper" may be designated as P^{0} or $P^{1.0}$.)

PMC = postmedial caesura (any emphatic MC-effect that occurs in an exposition *after* the first MC; a "second" MC-production, sounded several measures past an initial, fully successful MC.)

P^{rf} = the specialized P-theme within a Type 4 sonata—sonata-rondo—that also functions as a recurring, refrain theme with "rondo character," often also displaying a characteristic refrain-theme structure.

R1 = the initial ritornello (Ritornello 1 or opening tutti) at the opening of a Type 5 sonata (concerto movement). Similarly, R2, R3, and R4 stand for the second, third, and fourth ritornellos (or tuttis), each of which also has a specialized function and role to play within a Type 5 sonata.

R1:\ = prefix indicating material within R1 of a Type 5 sonata (concerto movement). (Thus R1:\P, R1:\S, and R1:\EEC

represent the modules functioning as the primary theme, the secondary theme, and the rhetorical EEC within the opening tutti of a Type 5 sonata.)

RT = retransition (a connective passage of preparation, usually leading to the onset of a new rotation, that is, to the repeat of the exposition, to the onset of the recapitulation, or to the beginning of the coda)

S = secondary-theme zone (follows an MC. This is built from precadential, pre-EEC thematic modules. Differing musical ideas within it, when they exist, are designated with superscripts as $S^{1.1}$, $S^{1.2}$, and so on. [See MMS and TMS.] A module that precedes or sets up the S-theme proper may be designated as S^{0} or $S^{1.0}$. Not to be confused with S1.)

S1 = the first solo section, Solo 1, of a Type 5 sonata (concerto movement), typically marked by the first entrance of the soloist following the orchestral R1 and ending with a trill cadence precipitating the onset of the second ritornello or tutti, R2. S1 is also the "solo exposition," even though, as discussed in Chapters 19 and 21, this is normally extended into a "larger exposition"—rotationally defined—with the addition of the immediately subsequent R2. Similarly, within concerto movements S2 and S3 stand for the second and third solo section. S2 is usually the developmental space of the Type 5 sonata. S3 (or sometimes R3⇒S3) is normally the "solo recapitulation," also extendable into the "larger recapitulation" with the addition of R4 (chapter 22). (Notice that in the concerto-space designation, S1, the numeral is not superscripted. When it is, as in S^{1}, S^{2}, and so on, it refers not to Solo 1 but to a portion of secondary-theme space. In concerto movements the two may appear in the same description, as with S1:\$S^{1.2}$, or "the second module of S-space within Solo 1 of a Type 5 sonata." See S.)

S1:\ = prefix indicating material within the S1 zone of a Type 5 sonata. (Thus S1:\P, S1:\S, and S1:\EEC represent the modules functioning as the primary theme, the secondary theme, and the EEC within the Solo 1 space of a Type 5 sonata. See S1.)

S^{C} = a theme within S-space (and thus before any clear articulation of an EEC) that, for any number of reasons, seems to take on the features and style more characteristic of a closing theme (C). Cf. $C^{pre\text{-}EEC}$.

TI = tutti interjection (in a Type 5 sonata, any brief, interrupting tutti impulse within what is otherwise a solo section, such as S1, S2, or S3. The first of these to appear, S1:\TI^{1}, shortly into Solo 1, is often formulaic and stylized, as noted in chapter 21.)

TMB = trimodular block (an especially emphatic type of multimodular structure in an exposition or recapitulation, always associated with the phenomenon of *apparent double medial caesuras*. Individual modules may be designated as TM^{1}, TM^{2}, and TM^{3}. Of these, TM^{1} and TM^{3} are usually "thematic." TM^{1}

follows the first apparent MC, TM^2 often reinvigorates the TR-style [often TM^1 merges into TM^2, $TM^1 \Rightarrow TM^2$] and helps to set up the second apparent MC, and TM^3 follows that second MC-effect. A TMB leads, at its end, to the EEC. Either TM^1 or TM^3 may give the impression of being the "real" S depending on the individual circumstances. Cf. TMS.)

TMS = trimodular S (a common type of MMS with three S-modules. Within the sonata narrative the first proves "unable" to produce a PAC; the second often thematizes the threat or difficulty; the third is a decisive cadential module. It differs from the TMB in its lack of apparent double medial caesuras: there is no second "apparent" MC after the second S module.)

TR = transition (following P, the energy-gaining modules driving toward the medial caesura)

TR⇒FS = the broad middle section of a continuous exposition that begins as a transition (TR) but at a crucial "point of conversion" midway through is often better described as *Fortspinnung* (FS) or, in other cases, a chain of thematic modules. Either procedure avoids producing a clear MC and the resultant two-part exposition. The ⇒ ("becomes" or "merges into") represents the conceptual point of conversion.

About the Authors

James Hepokoski, professor of music at Yale University, specializes in formal structure and hermeneutic issues in sonata-form-based repertories, ca. 1750–1920. He is the author of four books and numerous articles in a variety of areas, including Italian opera (Verdi), early-modernist composers (Sibelius, Strauss, Elgar), American and Germanic music-historical methodologies, and current literary-critical/cultural approaches to music. He was a co-editor of the journal *19th-Century Music* from 1992 to 2006.

Warren Darcy is professor of music theory and former director of the Division of Music Theory at the Oberlin College Conservatory. He has lectured and published widely on Wagnerian opera, and his book *Wagner's "Das Rheingold"* (Oxford, 1993) won the Society for Music Theory's 1995 Wallace Berry Award. He has also published on the music of Bruckner and Mahler and is currently engaged in a large-scale study of rotational form in Mahler's symphonies.

This page intentionally left blank

ELEMENTS *of* SONATA THEORY

This page intentionally left blank

CHAPTER ONE

Contexts

Differing Approaches to Sonata Form

There is no consensus regarding the manner in which sonata form in the decades around 1800 is to be grasped. On the contrary, analysts are confronted with a clutch of diverse approaches with differing emphases, interests, and terminologies. This is contested terrain, particularly since the structure is basic to how we conceptualize the Austro-Germanic art-music enterprise stemming from Haydn, Mozart, Beethoven, and Schubert. Our contribution, Sonata Theory, provides a *via media* among these approaches, remaining open to the positive insights that each has to offer and for the most part remaining methodologically compatible with them all. At the same time we propose new, genre-based perspectives, along with useful ways of formulating analytical questions and moving on to productive hermeneutic endeavors—interpretations of meaning.

Situating oneself within a conflicted field is a risky, fallible enterprise, in part because one is obliged reductively to characterize the work of others—and those others nearly always object (often rightly so) to such characterizations. And yet it may be helpful to sketch out some rough descriptions of viable approaches to the subject of sonata form, if only to suggest an impression of the larger playing field. In influential English-language scholarship today one might recognize four general trends: two broad musicological lines and two broad music-theory lines. To be sure, the categories overlap—they are anything but airtight—and within each there are differences and varied accents in the way the general method is formulated. Still, musicology and music theory have often pursued distinct paths, generating different questions and answers.

The two broad musicological approaches, sometimes intersecting, are: (1) the style of eclectic analytical writing favored by Donald Francis Tovey and carried on (and varied) by such differing writers as Joseph Kerman and Charles Rosen and (2) the more strictly "historical-evidentiary-empirical" concerns of such diverse figures as William S. Newman, Jan LaRue, Eugene K. Wolf, Leonard G. Ratner, and their successors. The two broad music-theoretical approaches are: (3) Schenkerian and post-Schenkerian methodologies and (4) lines of analysis emphasizing motivic growth from small musical cells, as well as the identification of phrase-shapes and the patterns of larger sectional blocks—a style of analysis associated with Arnold Schoenberg, Rudolph Réti, and Hans Keller, and including the work of Erwin Ratz and, most recently, William E.

Caplin.[1] At the risk of oversimplification (and with apologies to those unmentioned), we might characterize the interests of these four categories by citing an example of an important text within each.

1. Our first-category illustration is Charles Rosen's *Sonata Forms* (1980, rev. 1988).[2] Drawing on the analytical and prose style of Tovey and grounded in a vast knowledge of the repertory, Rosen's *magnum opus* stressed the variety of procedures that one can encounter in the "texture" or "process" that we now call sonata form. (Hence his plural, "forms," echoing Tovey.)[3] Rather than elaborating an intricate background plan for the form, Rosen preferred to demonstrate how difficult—or futile—it is to provide a set of detailed expectations regarding it because of the unique things that occur in individual pieces by composers of genius. As a matter of principle Rosen shunned the idea of a "general practice" for the construction of sonatas—except for a few tonal requirements and common textural choices—although there were clearly better and more masterly solutions to the general set of problems at hand.[4] This somewhat intuitive approach, acute and invariably musical, also emphasized the concept of tonal "polarization" (usually tonic and dominant) in expositions and famously regarded the expositional shift to a non-tonic key as an "opposition[al]" move, a "large-scale dissonance" ("structural dissonance" or "dissonant section") that needs to be resolved in the recapitulation.[5] A central feature of Rosen's writing (as well as that of Tovey and Kerman) was the description of individual compositional styles and preferences, along with the pronouncement of cleanly-divided aesthetic judgments of the works at hand—strong praise for the masterworks contrasted with tart dismissals of works deemed not to make the grade.

2. The second category is best represented by Leonard G. Ratner's *Classic Music* (1980).[6] Somewhat parallel to the scholarly-inventory work of William S. Newman and Jan LaRue, Ratner sought to reconstruct the concept of the eighteenth-century style from the point of view of the eighteenth century itself. The book was to be

> a full-scale explication of the stylistic premises of classic music, a guide to the principles according to which this music was composed. . . . The exposition of 18th-century musical rhetoric is found in theoretical and critical treatises. . . . [These writings] point to what was current *then*, illuminating our present view of the music. Coordinated with analysis of the music itself, the data gleaned from these writings make it possible to determine the basic criteria of expression, rhetoric, structure, performance, and style that govern classic music. . . . This book allows the student to approach the music and musical precepts of the 18th century in much the same way a listener of that time would have done.[7]

---

1. But even these broad categories are too limiting. Intermixed throughout them all are the various traditions passed on in the *Formenlehre*, the academic textbooks of form, which seem to have a separate reception-life of their own. In addition, other influential European perspectives that sometimes escape from or provide alternative havens within the above four categories have also proven provocative for current work—one thinks, for example, of the work of Jens Peter Larsen and Carl Dahlhaus. Moreover, in recent years differing scholars have begun to seek new ways to blend together formerly differing methodologies.

2. Rosen, *Sonata Forms*, rev. ed. (New York: Norton, 1988 [first ed. 1980]).

3. Donald Francis Tovey, "Sonata Forms," originally two different entries for the 11th (1911) and 14th (1929) eds., the latter of which is reprinted in Tovey, *Musical Articles from the Encyclopaedia Britannica* (London: Oxford University Press, 1944) [reissued in 1956 under the title *The Forms of Music*], pp. 208–32.

4. Rosen, *Sonata Forms*, rev. ed., pp. 4–7. Cf. the differing impression conveyed in W. Dean Sutcliffe's review, in *Music & Letters* 79 (1998), 601–4, of Rosen's modest revision of his earlier work *The Classical Style: Haydn, Mozart, Beethoven*, exp. ed. (New York: Norton, 1997 [orig. ed., 1971]). This review, in part, calls attention to the earlier book's apparent "emphasis on the normative aspects of the style . . . stereotypes and formulas"—concerns that raise a host of questions in these more skeptical times and ones that Rosen himself had sought to clarify in the later *Sonata Forms*.

5. Rosen, *Sonata Forms*, rev. ed., pp. 98–99, 229, 287. See also Rosen, *The Classical Style,* exp. ed., p. 33.

6. Ratner, *Classic Music: Expression, Form, and Style* (New York: Schirmer, 1980).

7. Ratner, Preface to *Classic Music*, pp. xiv–xvi.

Not surprisingly, Ratner paid close attention to the early theorists' descriptions of what came to be called (c. 1824–1845) "sonata form." The Newman-LaRue-Ratner projects (however they might differ in other respects) were ones of data-gathering and recovery. One of their features was to urge analysts to sideline nineteenth- or twentieth-century views of sonata form in order to gain a more period-conscious conception of the form.[8] (In this regard these interests are not without parallel to the performance-practice movement and its quest for "authenticity.") To varying degrees scholars within this circle seek to describe sonata form (and other forms) from the perspective of late-eighteenth-century theorists—favoring their terminology and concerns and being cautious about going beyond them.[9] Writers influenced by this point of view call upon the authority of late-eighteenth-century or early nineteenth-century writers on the form (such as the important statements of Heinrich Christoph Koch, Francesco Galeazzi, Augustus Kollmann, and Anton Reicha). Several of them have also tended to view harmony (modulations, key-areas visited, and so on) as the primary feature of sonata form in the years from roughly 1750 to 1820—giving it the upper hand over thematic arrangement. In the mid-twentieth century Ratner famously contested the earlier, thematic view of the sonata, which he regarded as discredited, an anachronistic, nineteenth-century (mis-)understanding of the form as it had been originally grasped in Beethovenian and pre-Beethovenian decades.[10] Some writers influenced by Ratner's work are also concerned with identifying historically defensible musical "topics" (standardized musical gestures or types within phrases) and eighteenth-century conceptions of "rhetoric" in this repertory.

3. Moving to the music-theory side of things, the touchstone of the third category is Heinrich Schenker's *Der freie Satz* (1935, translated as *Free Composition*).[11] For many music theorists interested in sonata form, no text is more central than this one. Opposed to traditional ways of discussing musical structure, Schenker was convinced that he had discovered a new theory of form, "a new concept, one inherent in the works of the great masters; indeed, it is the very secret and source of their being: the concept of organic coherence."[12] This theory was to be grounded not in phrase- or section-repetitions or in thematic manipulation but rather in linear-contrapuntal views of the sonata as the unfolding of a "fundamental structure" (*Ursatz*) by means of more elaborate middleground and foreground structures. Middlegrounds and foregrounds are understood as florid "diminutions" of more simple, elemental background gestures elaborated over the course of an entire movement. The method is highly sensitive to contrapuntal, linear voice-leading, long-range prolongations or descents of important individual pitches, and the like. Here sonata form is understood as divided into two parts (exposition-development || recapitulation) with a crucial harmonic "interruption" (||) at the end of the development and a subsequent rebeginning at the onset of the recapitulation, which restates and finally completes the fundamental structure interrupted at the end of the first part.[13]

---

8. See, e.g., Eugene K. Wolf, "Sonata Form," in *The New Harvard Dictionary of Music,* ed. Don Michael Randel (Cambridge, Mass.: Harvard University Press, 1986), pp. 764–67. This essay outlines the rhetorical-tonal structure at hand and provides a historical overview of the origins and transformations of the form.

9. In other respects Ratner-related styles of analysis seem to be musicological variants of the well-established sector of music theory, "history of music theory." A more purely music-theoretical analogue is Joel Lester, *Compositional Theory in the Eighteenth-Century* (Cambridge, Mass.: Harvard University Press, 1992).

10. The *locus classicus* of this position is Ratner, "Harmonic Aspects of Classic Form," *Journal of the American Musicological Society* 11 (1949), 159–68.

11. Schenker, *Free Composition* (German original, *Der freie Satz*, 1935), trans. and ed. (with additional commentary) Ernst Oster (New York: Longman, 1979). Especially relevant is part 3, ch. 5 ("Form"), pp. 128–145. "Section 3," on "Sonata Form" (including Oster's famous footnote), is found on pp. 133–41.

12. Schenker, *Free Composition*, p. xxi.

13. Also to be noted in terms of Schenkerian and post-Schenkerian analysis is the summary of sonata form in Allen Cadwallader and David Gagné, *Analysis of Tonal Music: A Schenkerian Approach* (New York: Oxford University Press, 1998), esp. ch. 11, "Sonata Principle," pp. 303–59. Similarly, one should mention William Rothstein, *Phrase Rhythm in Tonal Music* (New York: Schirmer, 1989), particularly ch. 4, "Phrase Rhythm

4. Our example of the fourth category is William E. Caplin's *Classical Form* (1998).[14] Its opening paragraph proclaimed the need for "a new theory of classical form," one that avoids "ill-defined concepts and ambiguous terminology derived from theories that have long fallen into disrepute." Following the work of Schoenberg and Ratz,[15] Caplin viewed form as a *grouping structure*, and he set out to identify and classify the "formal functions" of smaller thematic/formal units. In practice, this entailed close attention to the structures and subparts of three fundamental theme types: the *sentence* (consisting, for Caplin, of presentation, continuation, and cadential functions; or basic idea [usually repeated, perhaps with variation] + fragmentation + cadence); the *period* (antecedent + consequent); and the *small ternary* (A–B–A'). Much attention was also given to the anatomy of numerous "*hybrids*" that mix aspects of the more standard theme types (as defined by the author). As the musical parts are assembled, they can take on "framing functions," "interthematic functions," "harmonic functions," "initiating functions," "continuation functions," and so on, often at more large-scale levels. One aim of analysis is to be able to recognize the theme types (and hybrids) and to place them into a larger functional system of interrelated parts. In the end, what was provided was an elaborate taxonomy of different kinds of phrase-and-section juxtapositions.

The War against the Textbooks

One prominent feature of the study of sonata form in recent decades—very much in the wake of Tovey's similar assertions[16]—has been the repeated declaration that the "textbook" view of sonata form is inadequate to deal with the actual musical structures at hand. At best, such a scheme represents a conformist trap that master-composers avoid falling into. In addition, the implication has sometimes been that to undertake any such "textbook" description of norms, however nuanced or sophisticated, is a mistaken enterprise. It is not difficult to find conventionalized avowals on these matters. Here is a strong version of the credo from Claudio Spies, excerpted from an essay in a book of *Brahms Studies* (1991):

> There is nothing new about "forms" with whose aid pieces of music are easily and lazily categorized or typified, tagged, pigeon-holed, and conveniently stored away without further—or even prior—hearing, and without further thought. We were all initiated into the non-mysterious stolidities of "form," particularly the most fictitious one of all, "Sonata Form." Nor is there, I hasten to add, anything new in the notion that such "forms"—and especially "Sonata Form"—*are* fictions to whose specifications and proclaimed norms very few pieces of music worth any further thought actually conform in any appreciable way. . . . It is almost as if Brahms had decided to compose [the *Tragic Overture*] as a potent rebuttal of notions propounded by the tenets of *Formenlehre*, although [it] is by no means unique among his works in this respect.

The same point, put more gently—and after an admirably detailed study of Brahms—may be found from James Webster in the same volume:

> From examples like these it is clear that norms of formal procedure, whether the bad old textbook

---

and Form: Some Preliminaries," pp. 102–20. This is an analytically sophisticated discussion of forms in general (including sonata form) and, in part, it seeks to blend some of the concerns of Schenkerians with the more musicological (and often emphatically non-Schenkerian) studies by Rosen, Ratner, and others.

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

15. Schoenberg, *Fundamentals of Musical Composition*, ed. Gerald Strang and Leonard Stein (London: Faber & Faber, 1967); Ratz, *Einführung in die musikalische Formenlehre: Über Formprinzipien in den Inventionen und Fugen J. S. Bachs und ihre Bedeutung für die Kompositionstechnik Beethovens*, 3rd ed., enl. (Vienna: Universal, 1973 [1st ed., 1951]).

16. See, e.g., Donald Francis Tovey, "Some Aspects of Beethoven's Art Forms" and "Musical Form and Matter," in *The Main Stream of Music and Other Essays* (Oxford: Oxford University Press, 1949), pp. 272–73, 160–62; and Tovey, "Sonata Forms," pp. 210–12 ("There are no rules whatever for the number or distribution of themes in sonata form").

> models or the numerical averages developed earlier in this paper, can never satisfactorily account for the reality of individual compositions. In fact, when Brahms's technique seems most paradoxical—as in the timeless, themeless, tonic retransition we have just analysed—the artistic result is often the most poetic.[17]

Remarks along these lines could hardly be more familiar. Even earlier, by midcentury, it had become a scholarly point of honor to declare war on the textbooks and, for some (again, in varying degrees), on the often-wooden limitations of classifying schemes in general. Whether uttered in stronger or gentler versions, such declarations advanced unswervingly orthodox late-twentieth-century convictions, and they were caught up in the traditional philosophical dilemma of universals and particulars. For the most part—again, much as Tovey had done—they took partisan positions on behalf of the particulars, or at least on behalf of the ultimate noncapturability of the great masters. Apart from assessing this neonominalist argument on its own terms, it would also be valuable to investigate the modernist assumptions that made such views possible: the mystification of genius; the belief in the compulsion of the true artist to escape from confining, externally applied rules or systems; the precept that what we most revere in music must not only be beyond the grasp of academic minds and rational classification but must always be declared to be so; and so on.

Studying and teaching musicology and music theory in the 1960s, 1970s, and 1980s, the authors of this book absorbed such views into the marrow of their bones. We also agreed—and we continue to agree—that prior textbooks had invited a too rigid understanding of sonata form. So far as the gravamen of the charge goes, the literal point is correct and has the added benefit of bringing caution to any new analytical inquiry. Still, the problem of determining the role of convention within this "classical" repertory was more complex than the reflex denunciations suggested. The reiterated conviction that there was no single plan for sonata form in the later eighteenth century, true enough in its narrow, literal sense, rises to the level of an error when it is naively taken either to dismiss the presence of substantially more complex systems of standard practices or to discourage inquiry into those practices. Is there a more effective way of examining conventional musical gestures (or calling forth that which *was* conventional within individualized musical gestures) without producing ideas that were reductive, stiff, mechanical, prescriptive? Is an aesthetically sensitive openness to the study of convention within composition possible?

The most strongly formulated arguments against generalized principles of sonata practice concealed a substantial weakness: in their intensity they tempted one to overstate the degree to which such classifications were ever intended to be equivalent to scientific laws. Within the humanities norms, generic options, and more-or-less standard procedures are not laws at all. And since they are not, there was no need to suppose that the existence of numerous exceptions or deviations invalidated the norm. Perhaps the many deviations were purposeful dialogues with the background norm. But this would mean, paradoxically, that the deviations helped to reinforce the socially shared norm that was being temporarily overridden. (Otherwise how could they be perceived as deviations at all?) But what is meant by a norm? And how could one come to an understanding of what such norms might have been? We began to seek a way out of the dilemma. The most profitable guidelines for our solution lay within the domains of current genre theory and hermeneutics.

Given the flexibility found in the large-scale architecture of later-eighteenth-century composition, the main descriptive problem was the difficulty of positing convincing categories of typical procedures. As scholars of eighteenth-

17. Spies, "'Form' and the *Tragic Overture*: An Adjuration," and James Webster, "The General and the Particular in Brahms's Later Sonata Forms," in *Brahms Studies: Analytical and Historical Perspectives*, ed. George S. Bozarth (Oxford: Clarendon, 1990), pp. 391 (Spies) and 75 (Webster).

century music perennially point out, surprising occurrences and variants abound—all the more so when one's investigation takes a panoramic view, extending beyond Haydn, Mozart, and early Beethoven to include the works of less-explored composers. It is for this reason that attempts to describe normative sonata procedures tend to bog down in trying to account for a host of seemingly unusual cases (of which there is an especially abundant supply in Haydn's *œuvre*).

So much is evident, but the only alternative to throwing up one's hands in the face of such diversity (rallying around the cry, "Anything can happen!," which is obviously untrue) was to find a reasonable middle ground between confiningly rigid schemata and the claim of a near-total freedom. It was necessary to retrieve a workable hermeneutic space between the reductive textbook models of the nineteenth and early twentieth centuries and the unhelpful (though still fashionable) "lowest-common-denominator" harmonic models, whose claims to adequacy have been challenged on both historical and conceptual grounds in an important essay from 1991 by Mark Evan Bonds.[18] In that essay Bonds distinguished between "conformational" and "generative" concepts of sonata form, traced the fortunes of these concepts historically, and submitted the mid-twentieth-century ascendancy of the generative models to a critique. Among his conclusions:

> Few analyses [today] openly acknowledge the extent to which composers worked within the context of formal conventions. . . . But it would be ludicrous to argue that sonata form was not at least in part an *a priori* schema available to the composer. . . . Sonata form, for Haydn, was in fact a point of departure, a mold, albeit a flexible one. . . . What is needed, then, is a general theory of form that can account for conventional patterns and at the same time do justice to the immense diversity that exists within the framework of these patterns.[19]

Thus the challenge: to articulate the implied pattern-types that appear in some of the clearest or most notable exemplars and to do this with as much detail and specificity as the material encourages. These heuristic norms need not be considered as literally existing "things." Rather, they may be understood as what Dahlhaus, following Max Weber, regarded as ideal types or what we prefer to consider as regulative guides for interpretation. Moreover, these norms would have to be defined neither by unusual cases nor by expressive deformations of more standard choices. Rather, they would derive from the standard choices themselves, insofar as the frequency of those choices (not their inevitability) permit one, inductively, to infer a background set of guidelines shared by composers and a community of listeners at a given historical time and place. As we constructed these models, then, we were concerned to identify types or tendencies that (in retrospect) were influential generic participants in the eventual crystallization or early reification phase of the sonata in the mid-eighteenth century, when the preferred options became both clearer and somewhat more consistent.[20] The result was the system that we call Sonata Theory.

Our intention is not to lay down binding laws or invariant rules concerning either the parts of a sonata or the sonata as a whole. Instead, we are trying to sketch the outlines of a complex set of common options or generic defaults. It is not that any attempt to recover standard pat-

---

18. Bonds," The Paradox of Musical Form," ch. 1 of *Wordless Rhetoric: Musical Form and the Metaphor of the Oration* (Cambridge, Mass.: Harvard University Press, 1991), pp. 13–52.

19. Bonds, "The Paradox of Musical Form," p. 29.

20. E.g., as articulated in Wolf, "Sonata Form," *The New Harvard Dictionary of Music*, p. 766: "By about 1765, however, full sonata form [i.e., with full recapitulation]—though never the rigid textbook variety—was rapidly becoming the norm in fast movements and many slow movements of symphonies and related genres, works for chamber ensemble, and solo and accompanied sonatas in all but a few major centers." Similar observations regarding the increasing normativity of certain kinds of sonata procedures—especially those identified with the Viennese Classicism of Haydn, Mozart, and early Beethoven in the period circa 1770–1800—may be found in the writings of virtually every author who has investigated such things. See, e.g., the many similar remarks in Charles Rosen, *Sonata Forms*, rev. ed., pp. 145, 153, 156–58, 161, and 286–87.

terns is a flawed enterprise; rather, it is that prior attempts have been inadequately conceived. We offer Sonata Theory as a heuristic construct that can help the task of analysis and hermeneutics. At any point, the method outlined here can be expanded or modified through criticism, correction, or nuance. Indeed, we invite this. The proposed construct is intended only as a beginning, as a work-in-progress—not as a fixed set of finalized dicta. As an assemblage of separate subparts, each of which should be subjected to constant testing and refinement, the utility of Sonata Theory as a whole does not rest on the unexceptionable validity of any correctible subpart.

Sonata Theory: Introductory Remarks

What follows lays the groundwork of a method of approaching analytically any sonata-form movement from the period of Haydn, Mozart, and Beethoven. A central premise of this method is the conviction that we must seek to understand the backdrop of normative procedures within the different zones or action-spaces of the late-eighteenth-century sonata. Much of this book sketches out key technical features of those norms as we currently understand them.

At any given point in the construction of a sonata form, a composer was faced with an array of common types of continuation-choices established by the limits of "expected" architecture found in (and generalized from) numerous generic precedents. (To produce a keyboard-sonata or symphonic movement was to place one's individual achievement into a dialogue with a community-shared pool of preexisting works, probably including some well-known ones, that formed the new work's context of understanding.) This is not to say that any skilled composer soberly pondered these choices, one by one, in the act of composing. Surely the most common decisions were made efficiently, expertly, and tacitly on the basis of norms that had been internalized (rendered automatic) through experience and familiarity with the style. Still, even before a sonata form was begun, a composer might, consciously or not, confront an array of initial questions acting as a filter for all that followed: symphony movement? overture? sonata? chamber music? how long or "grand" a movement? how complex? how "original"? how "intense" or "challenging" to listeners? what is the expected audience? for connoisseurs or amateurs (*Kenner* or *Liebhaber*)? how "unusual" in its internal language and manner of presentation? in competition with whom? whom am I trying to impress? for what occasion? and so on.

Once these gateways had been determined and work begun in earnest—the task of creating an engaging musical pathway through pre-established, generically obligatory stations—the composer faced practical issues of musical continuation from one idea to its successor. (A succeeding phrase, even an utterly contrasting phrase, would typically be heard as "reacting to" what had been established up to that point—moving outward to another branch of the musical ramification.) A sonata form required that certain audible goals be successively articulated and secured, even though the individual details of each sonata journey could differ remarkably. A composer's choices involved not only varying senses of the propriety of "what sorts of things could reasonably be expected next" within the style but also how delectable surprises, even varying degrees of seeming transgressions, might be folded into the expanding network of ideas. Within each compositional zone (action-space) or subsection these "internalized" features included such things as generically appropriate types of themes and textures; reasonable lengths of individual passages (which depended on the anticipated length and complexity of the whole composition); dynamics; degrees of anticipated contrast; standard "topics" or thematic formulas; properly placed cadences and/or cadential delay or frustration; the handling of major- and minor-mode coloration; boundaries of taste; and the limits of eccentricity.

The options available from compositional zone to zone existed conceptually within the knowledgeable musical community as something on the order of tasteful generic advice—enabling and constraining guidelines (not inviolable rules) within the "sonata-game"—given by a shared knowledge of precedents. Moreover, the available guidelines for each moment (pri-

mary theme, transition, medial caesura, secondary theme, and so on) were not accessible in an arbitrary, non-weighted fashion. Some choices were virtually obligatory; others less so, sometimes in discernible degrees. (For novice-composers, one might wittily fantasize—provided that the image is not taken too literally—something on the order of an aggressively complex "wizard" help feature within a late-eighteenth-century musical computer application, prompting the still-puzzled apprentice with a welter of numerous, successive dialog boxes of general information, tips, pre-selected weighted options, and strong, generically normative suggestions as the act of composition proceeded. What would have been urged here were such things as thematic-modular shape, style, effect, and format appropriate to the relevant action-space moment—not literal content, the burden of which was still placed on the composer.)

Within the late-eighteenth-century style some of the options were much more frequently chosen: To suggest the strength and pre-established hierarchical ordering of these options we call the more normative procedures *first- and second-level defaults* within the various zones.[21] Most simply put, composers selected (or adapted) first-level options more frequently than second-level ones, and so on. (Writers of minor-mode sonatas, for instance, more often modulated to the major mediant, III, in the exposition, than to the minor dominant, v—a less common option.) As we use it, however, the term *default* connotes more than a merely preferred option for otherwise detached consideration. First-level defaults were almost reflexive choices—the things that most composers might do as a matter of course, the first option that would normally occur to them. More than that: not to activate a first-level-default option (for example, to provide an expositional move to v instead of to III) would require a more fully conscious decision—the striving for an effect different from that provided by the usual choice. An additional implication is that not to choose the first-level default would in most cases lead one to consider what the second-level default was—the next most obvious choice. If that, too, were rejected, then one was next invited to consider the third-level default (if it existed), and so on. Or perhaps at some point in this process a composer might decide to do something unusual by rejecting all of the default choices altogether, in pursuit of a *deformation* of that compositional moment.

As might be imagined, the whole system was highly complex, typically involving at any compositional point more than two default levels of options. This is why it requires so much time—and space—to reconstruct the background system. But it is only through an understanding of what the main options were that we can come to grips with the implications of a composer's choices from moment to moment.

In confronting any individual composition we seek to determine which gestures in it were normative within the style, which were elaborate, elegant, or strained treatments of the culturally available norms, and which were not normative at all. Sonata Theory starts from the premise that an individual composition is a musical utterance that is set (by the composer) into a dialogue with implied norms. This is an understanding of formal procedures as *dynamic, dialogic*. Our conception of the sonata as an instance of *dialogic form* is not accurately described as seeking to reinstate a bluntly "conformational" view of that structure (in Bonds's original sense of that category). Viewed more subtly, it is not the obligation of a sonata to "conform" to a fixed background pattern, which then, in turn, might be construed as an "ideal" or "well formed" shape from which deviations might be regarded as compositional errors or aesthetically undesirable distortions. Rather, the composer generates a sonata—which we regard as a *process*, a linear series of compositional choices—to enter into a dialogue with an intricate web of interrelated norms as an ongoing action in time. The acoustic surface of any sonata form (what we literally hear) sets forth

21. At some level the literal, computer-definition concept of *default*—an assumption prebuilt into the large-scale automatic (but alterable) decisions of a software program at the moment of its initialization—is not fully congruent with our free adaptation of it here (in the sense of ongoing, strongly weighted advice, standard choices, and normatively arrayed options). As mentioned earlier, the metaphorical implication, if applicable at all, is to be worn loosely.

the sonic traces of this individualized, processual dialogue, one that, from the standpoint of reception, it is the task of the analyst to reinvigorate. The backdrop of norms against which a sonata or any of its successive zones is placed into dialogue is no monodimensional, reified "thing." On the contrary, that backdrop comprises complex sets (or constellations) of flexible action-options, devised to facilitate the dialogue. Understanding form as dialogue also helps us to realize that in some cases standard procedures may be locally overridden for certain expressive effects. These effects differ from composition to composition: each needs to be interpreted individually. The more piece-specific one's readings can be along these lines, the better. In any analysis merely to assert that something is done "for expressive reasons" or "for reasons of variety" is obviously inadequate.

Background norms and standard options are classifiable into common and less common selections at different times and different places. Within an individual composition, a markedly exceptional procedure here or there is just that—exceptional. We call such an occurrence a generic deformation: a stretching or distortion of a norm beyond its understood limits; a pointed overriding of a standard option. The term "deformation," in this specific context, is a narrow-definitional, technical one, grounded in precedents in literary theory and other research areas. In its strictly limited, analytical usage within Sonata Theory, "deformation" carries no negative charge, no negative assessment. On the contrary, such deformations are typically engaging, aesthetically positive occurrences that contribute to the appeal and interest of a piece. As we use the term, it signifies only a purposely strained or non-normative realization of a musical action-space, a surprising or innovative departure from the constellation of habitual practices, an imaginative teasing or thwarting, sometimes playful, of expectations, presumably in order to generate an enhanced or astonishing poetic effect.[22]

Deformations—unusual or strongly characterized, *ad hoc* moments—are common within the works of many different late-eighteenth-century composers. Indeed, they are rampant in Haydn, who delighted in producing surprising effects. Such occurrences, in dialogue with a norm, should not be regarded as redefining that norm unless the composer continued to employ that idiosyncratic feature in other works (thus customizing the norm for his own use) or unless later composers picked up the deformation as one of their more or less standard options. When this later occurrence happens, the original exception is no longer to be regarded as a deformation *per se* but becomes one of the lower-level defaults within the Sonata-Theory system. What was a deformation in Beethoven could become a lower-level default in Schumann, Liszt, or Wagner—part of a larger network of nineteenth-century sonata-deformation families.

The essence of Sonata Theory lies in uncovering and interpreting the dialogue of an individual piece with the background set of norms. This style of analysis considers every aspect of the individual work: themes, harmonic and contrapuntal motion, large- and small-scale shapes, textures, dynamics, instrumentation, tempos, repeat conventions, and so on. The main requirement for the application of the method is to grasp the controlled flexibility of the implicit underlying system of conventions. Elaborating that system is the goal of the *Elements*.

At every turn, our aim has been to focus on the most basic features of the sonata and never to forget why we perform and listen to this music in the first place. To overlook fundamental things leads one's analyses astray or renders them sterile, bookish, or irrelevant. The best analytical system is the one that seeks to reawaken or

22. It would be a mistake, therefore, to read into this technical usage any residual connotations of the evaluatively negative, such as the "*de*formed" (in its more typical meaning), the "*dis*figured," the "*miss*hapen," the "*ab*normal," the "*poorly* formed," or the "ugly." Those are not our connotations, and within the framework of Sonata Theory terminology we distance ourselves from them as strongly as we can. The central thing is to be able to grasp the intended nuance of the technical term "deformation"—to be able to perceive in it a genre-enabled, *positive* sense of strain, a deliberately manufactured tension set apart in this aesthetic-analytical, "artificial" context from any implication of criticism or (much less) censure. These connotational points are revisited and amplified in the "Deformation" section of appendix 2, which also offers further reflections on the concepts of *dialogic form* and sonata-form action-spaces.

re-energize the latent drama, power, wit, and wonder within individual compositions. Whenever an analytical system diverts attention from the impact of the music as real experience—or, even more, when it fails to heighten our own experience of the music—then that analytical system is in need of correction. We hope that Sonata Theory, in its practical application, will lead beyond the academic explanations and interpretations of the self-enclosed work into a larger reflection on the changing meanings of this music within society.

In part, we do this by redirecting analytical attention to those portions of the sonata that have been taken for granted or passed over in relative silence in most preceding discussions. These include the composer's treatment of *caesuras* (medial and final), the textural drive toward important *cadences* (including especially the moments of what we call *essential expositional closure [EEC]* and *essential structural closure [ESC]*), the *rotational* aspect of the sonata movement as a whole (its tendency to cycle repeatedly through large, thematically differentiated blocks), and many other considerations. Although this was by no means clear to us when we began this project, one result of our work has been to defamiliarize the sonatas of Haydn, Mozart, and Beethoven—permitting us to hear them in what we have found to be more rewarding ways. To some extent, we discovered early on that we often had to overcome our own patterns of habituation in analysis and understanding "in order [to adapt the words of Viktor Shklovsky] to return sensation to our limbs, in order to make us feel objects, to make a stone feel stony."[23] The idiosyncratic concerns—even the idiosyncratic terms—of Sonata Theory can help in this regard.

For the authors, one of the most challenging burdens in devising Sonata Theory has been to remain willing to submit all components of currently "orthodox wisdom" regarding sonatas to radical questioning—comfortable trenches of thought that had long been part of our own reflexive modes of approaching this music. From the beginning we sought to listen carefully to this repertory, trying to remain open to what it seemed to want to tell us on its own terms, insofar as we could apprehend those terms in our own, very different times. Before long we came to understand that everything that we had considered to be established about sonata-analysis had to be rethought. If only for this reason, we realize how curious Sonata Theory might at first appear, especially to scholars habituated within other modes of analysis and accustomed to other kinds of theoretical questions. The value of any analytical system, however, lies in the robustness of its interpretive power. It is that interpretive adequacy that we have been seeking. Whenever existing terminology was adequate, we have retained it; whenever it was misleading or connotatively unhelpful, we have decided to change it; whenever it lacked a term for a crucial concept, we have been obliged to devise a new one.

Readers might initially find that the basic concerns of Sonata Theory are learned relatively quickly—like the moves of chess. These concerns may seem simple precisely because they are simple. At all points in the analysis of a sonata, we have tried to emphasize the most essential features and dramatized musical goals. Beyond the elementary principles of Sonata Theory, though, lies an elaborate network of possibility, nuance, flexibility, sophistication, and detail that takes patience to master. As with chess, again, one may learn the moves rapidly, but to play the game at a fully proficient level is more difficult. Notwithstanding its many postulates and axioms, Sonata Theory is no mechanical system. Rather, in proper application it is an art that requires training, musical sensi-

23. Shklovsky, *Theory of Prose* [from second edition, 1929], transl. Benjamin Sher (Elmwood Park, Ill.: Dalkey Archive Press, 1990), p. 6. In order to accomplish these things, declared the Russian Formalist Shklovksy, "[we have] been given the tool of art. . . . By 'enstranging' objects and complicating form, the device of art makes perception long and 'laborious'. . . . *Art is a means of experiencing the process of creativity. The artifact itself is quite unimportant*." Sher defends his translation, "enstranging" (as opposed to the more traditional choices, "defamiliarizing" and "estranging"), on p. xix.

tivity, and much experience with the repertory in question.

At the heart of the theory is the recognition and interpretation of *expressive/dramatic trajectories toward generically obligatory cadences*. For the present, we might only register the degree to which this concern resonates with Heinrich Schenker's much-quoted description of musical motion and dramatized process in *Free Composition* (*Der freie Satz*, 1935):

> The *goal* and the course to the goal are primary. Content comes afterward: without a goal there can be no content.
>
> In the art of music, as in life, motion toward the goal encounters obstacles, reverses, disappointments, and involves great distances, detours, expansions, interpolations, and, in short, retardations of all kinds. Therein lies the source of all artistic delaying, from which the creative mind can derive content that is ever new.[24]

24. Schenker, *Free Composition*, p. 5.

CHAPTER TWO

Sonata Form as a Whole

Foundational Considerations

Sonata form is the most important large structure of individual movements from the "common-practice" tonal era. It sets forth and resolves its musical discourse within a large-scale binary format. The term "sonata form" was almost surely unknown to Haydn, Mozart, early Beethoven, and their contemporaries: it seems to have surfaced only in the 1820s and 1830s. In the late-eighteenth and early-nineteenth century this structure would have been grasped primarily as the customary design of first movements within sonatas, chamber music, and symphonies, although it was by no means confined only to first movements (nor only to rapid-tempo movements). The varying descriptions from contemporary theorists were more convoluted. There the form was variously described as: "the first allegro of the symphony [or sonata]" disposed in "two sections" [*zwey Theile*] and three "main periods" [*Hauptperioden*] (Koch 1793); within "larger pieces of music" a "well-conducted melody [!] . . . divided into two parts, either connected, or separated in the middle by a repeat sign" (Galeazzi 1796); "an elaborate movement [or] a long movement . . . generally divided into *two sections*" (Kollmann 1799); "grand binary form" [*grande coupe binaire*] (Reicha 1826); and so on.[1] Still, "sonata form" (*Sonatenform*) seems to have been a familiar term by the mid-1820s, at least in A. B. Marx's *Berliner allgemeine musikalische Zeitung* circle, where it referred both to the multimovement cycle as a whole and, occasionally, to the form of an individual movement.[2] It was only in 1838 and 1845, though, in technical

1. Heinrich Christoph Koch, *Versuch einer Anleitung zur Composition* (Leipzig: Adam Friedrich Böhme, 1793; rpt., Hildesheim: Georg Olms, 1969), pp. 304–5 (from section 101), trans. Nancy Kovaleff Baker in Koch, *Introductory Essay on Composition: The Mechanical Rules of Melody, Sections 3 and 4* (New Haven, Conn.: Yale University Press, 1983), p. 199; Francesco Galeazzi, *Elementi teorico-pratici di musica*, vol. 2 (Rome: Puccinelli, 1796), the relevant extracts of which were excerpted and translated in Bathia Churgin, "Francesco Galeazzi's Description (1796) of Sonata Form," *Journal of the American Musicological Society* 21 (1968), 181–99 (above quotations from pp. 189–90); A. F. C. Kollmann, *An Essay on Practical Musical Composition* (London, 1799; rpt. New York: Da Capo Press, 1973), p. 4 [ch. 1, section 10]; Anton Reicha, *Traité de haute composition musicale* (Paris, 1826), discussed, e.g., in Ian Bent and William Drabkin, *Analysis* (New York: Norton, 1987), pp. 18–20, and especially Peter A. Hoyt, "The Concept of *développement* in the Early Nineteenth Century," in *Music Theory in the Age of Romanticism*, ed. Ian Bent (Cambridge: Cambridge University Press, 1996), pp. 141–62.

2. In the journal's first year of publication (1824) the term 'sonata form" appeared in both senses. The first, apparently initially the more common, was a description of the entire multimovement cycle (used by Marx,

discussions of the form's particulars, that Marx put the stamp of approval on the term "Sonatenform" with regard to the individual-movement structure.[3] Throughout this book we use that term as a familiar quick-reference, even as we realize that that designation was not current in the eighteenth century.

Sonata form is neither a set of "textbook" rules nor a fixed scheme. Rather, it is a constellation of normative and optional procedures that are flexible in their realization—a field of enabling and constraining guidelines applied in the production and interpretation of a familiar compositional shape. Existing at any given moment, synchronically, as a mappable constellation (although displaying variants from one location to another, from one composer to another), the genre was subjected to ongoing diachronic transformation in history, changing via incremental nuances from decade to decade. Haydn's conception of what was customary within sonata form in 1770 differed somewhat from Beethoven's conception in 1805. However such models might be said to have differed, they also shared certain crucial, genre-defining features that make them all recognizable as sonata form. Here we are dealing primarily with the model that crystallized during the second half of the eighteenth century and that reached a peak in the mature works of Haydn and Mozart and the early works of Beethoven.

What we now call sonata form was developed as a response to aspects of the world view of the Enlightenment and the concomitantly emerging modernism. Considered generally, it could be understood as an abstract metaphor for disciplined, balanced action in the world, a generalized action involving differing types of idealized mid- and late-eighteenth-century personalities. (Its potential for "extramusical" connotations and analogues is discussed in the final section of chapter 11, "Narrative Implications: The Sonata as Metaphor for Human Action.") Sonata form emphasized short-range topical flexibility, grace, and forward-driving dynamism combined—in both the short and long range—with balance, symmetry, closure, and the rational resolution of tensions. By the mid-eighteenth century it had become obligatory for the first movement of a standard multimovement instrumental work; it had also become a common, if optional, choice for the slow movement and the finale. Slow movements and finales sometimes also displayed different adaptations of the form. Although the guidelines in most of this book were written predominantly with first and last movements and single-movement overtures in mind (all energetic "Allegro movements"), they are also applicable, occasionally with some modifications, to slow movements.

From the compositional point of view sonata form was an ordered system of generically available options permitting the spanning of ever larger expanses of time. A sonata-form project was a feat of engineering, like the construction of a bridge "thrown out" into space. In the eighteenth-century style this temporal span was to be built from rather simple materials: trim, elementary musical modules whose brevity and small-scale balances seemed best suited to short-winded compositions. In the hands of most composers, constructing a sonata-form movement was a task of *modular assembly*: the forging of a succession of short, section-specific

---

Heinrich Joseph Birnbach, and others), a usage that persisted throughout much of the nineteenth century, especially in German-speaking regions. The other use of "Sonatenform" referred to the structure of an individual movement. It first appeared in a casual, unexplained way—as if it were already a common label—in Marx's 1824 essay on the E-minor second movement (Prestissimo) of Beethoven's Piano Sonata in E, op. 109 ("Es bildet mit dem letzen Satze die eigentliche Sonate und ist auch in der Sonaten-Form hingeworfen," *BamZ*, I, 1824, 37b) and in Carl Loewe's discussion of the first movement of Beethoven's Cello Sonata, op. 102 no.1 ("Hart und rauh, im männlichen Zorne, beginnt ein kurzes Allegro (A-moll) in der Sonatenform," *BamZ*, 1824, 410b). See the discussion of terminology and quotation of sources in the entry by Hans-Joachim Hinrichsen, "Sonatenform, Sonatenhauptsatzform" [1996], in Hans Heinrich Eggebrecht, ed., *Handwörterbuch der musikalischen Terminologie* (Stuttgart: Steiner, n.d.), pp. 1–7.

3. A. B. Marx, *Die Lehre von der musikalischen Komposition, praktisch-theoretisch*, vols. 2 and 3, 1st eds. (Leipzig, 1838 and 1845), 2:482, 497; 3:195; cited in Hinrichsen, "Sonatenform, Sonatenhauptsatzform" [1996], pp. 6–7. See also Marx, *Musical Form in the Age of Beethoven: Selected Writings on Theory and Method*, ed. and trans. Scott Burnham (Cambridge: Cambridge University Press, 1997).

musical units (spaces of action) linked together into an ongoing linear chain—pressing down and connecting one appropriately stylized musical tile after another.[4] One of the challenges facing the mid- and late-eighteenth-century composer was to use a seemingly unassuming, *galant* language, grounded in structural punctuation and periodicity, to produce ever more spectacular spans for occasions of enhanced dignity, prestige, or social importance. Ever-larger, thematically differentiated binary structures (sonata forms, often with built-in repetitions of individual sections), eventual accretions to the structure (slow introductions and longer codas), and multimovement conventions all had their roles to play in this process of generic enlargement. And ultimately they led to the grandly monumental, personalized structures of Haydn, Mozart, and Beethoven.

The most typical sonata forms (what we call Type 3 sonatas) articulate an overall rounded binary structure. The two parts of this larger structure are, in modern terminology: (1) the exposition and (2) the development and recapitulation. As will be elaborated at the end of this chapter, both parts may be marked for repeat, or the composer may eliminate the repeat of part 2 or, under some circumstances, both repeats. Notwithstanding its binary origins, the normative, Type 3 sonata consists of three musical action-spaces (again, the exposition, development, and recapitulation), laid out in a large A||BA' format. Hence the common observation that the form consists of an originally binary structure often arrayed in a ternary plan. Each of the three spaces is usually subjected to thematic and textural differentiation. Each is marked by several successive themes and textures, all of which are normally recognizable as generically appropriate for their specified location. These three spaces can be viewed as expansions of the three phases of the continuous rounded binary form (the rounded binary structure in which the first part ends in a secondary key). We shall take up these spaces individually. (In figure 2.1 we have provided two diagrams of Sonata Theory's conception of the most common type of sonata form: 2.1a refers to the exposition; 2.1b to the whole sonata-form movement.)

Exposition

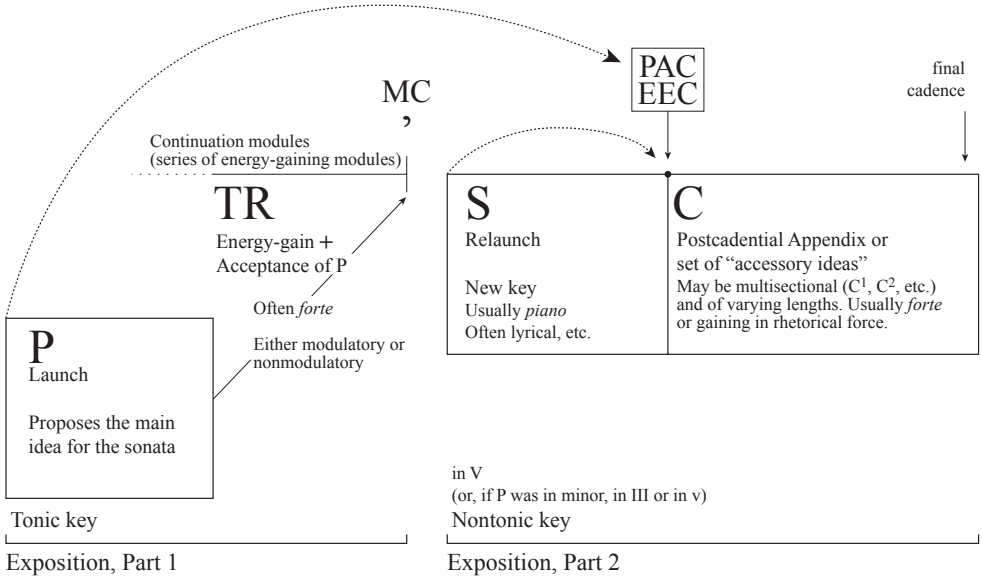
As with all of the action-spaces the exposition is assigned a double-task, one harmonic and the other thematic-textural ("rhetorical"). Its *harmonic* task is to propose the initial tonic and then, following any number of normative (and dramatized) textural paths, to move to and cadence in a secondary key. In major-mode sonatas—the most common in the eighteenth century—this was the key of the dominant (which may be indicated as V_T, meaning "a V that is tonicized"), thereby generating tonal tension. In minor-mode sonatas this was usually the key of the major mediant (III), although a less-often-selected choice (second-level default) was the minor dominant (v). The differing psychological and structural world of minor-mode sonatas is dealt with in chapter 14. Here, for the most part, we shall focus on major-mode practice.

The exposition's *rhetorical* task, no less important, is to provide a referential arrangement or layout of specialized themes and textures against which the events of the two subsequent spaces—development and recapitulation—are to be measured and understood. We refer to this layout as *Rotation 1* or the *expositional rotation*.[5]

4. To be sure—and particularly in the hands of the master composers of the period—certain passages within individual sonata forms may from time to time give the impression of a broader continuity of internal ramification. This is especially the case with the startlingly original musical language of Haydn, who, even within a generally modular and "sectionalized" concept of formal practice, often favored passages of ongoing *Fortspinnung* (a moment-to-moment "spinning-out" of modular growth and elaboration). For brief characterizations of Haydn's often-"vitalistic" compositional style, see ch. 11, subsection "Recompositions, Reorderings, Interpolations" (especially n. 2 and the text to which it refers), and ch. 18, subsection "Haydn's Treatments of Type 4 Finales" (especially n. 49 and related text).

5. Sonata-form structures are centrally concerned with the formal principle that we call *rotational form* or the *rotational process*: two or more (varied) cyclings—*rotations*—through a modular pattern or succession laid down at the outset of the structure. Appendix 2 provides a broader introduction to this principle, which pervades the discussion of sonata form in this book.

a. Exposition only: the Essential Expositional Trajectory (to the EEC)



b. The entire structure: the Essential Sonata Trajectory (to the ESC)

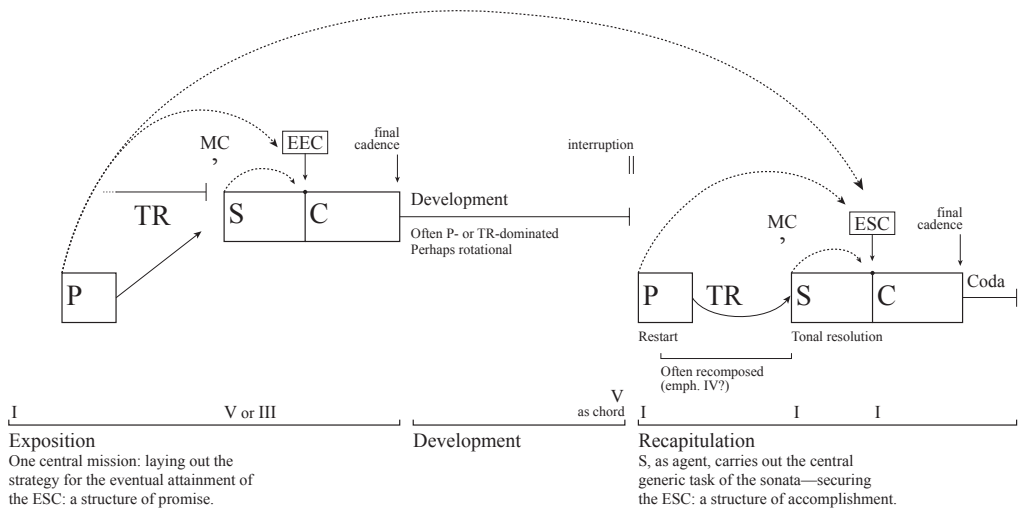


FIGURE 2.1 The Generic Layout of Sonata Form

Because the exposition's succession of events serves, especially in its second half, to predict the plan and purpose of the entire third space—the recapitulation, which finally resolves the work—its layout may be understood as articulating a *structure of promise* (indicating how it proposes that "things work out" in the recapitulatory rotation-to-come). Because the arrangement of rhetorical modules in Rotation 1 provides the ordered set of events that articulates the uniqueness and specific personality of that piece, it should be kept in mind when assessing all of the later events in the movement.

Within the expositional rotation the tonal and rhetorical tasks unfold simultaneously, intertwined with each other in mutually reinforcing ways. The exposition begins with a *primary theme* or *primary idea* (P) in the tonic that sets the emotional tone of the whole work. The most common layout for the remainder of the exposition continues with an energy-gaining zone of transition (TR) that leads to a mid-expositional break or medial caesura (MC). This is typically followed by the onset of a specialized, secondary-theme zone (S) in the new key. The generically essential tonal purpose of the exposition is to drive to and produce a secure perfect authentic cadence (PAC) in the new key (notated as V:PAC in major-mode sonatas, III:PAC or v:PAC in minor-mode ones). We refer to the first satisfactory PAC within the secondary key that goes on to differing material as the point of *essential expositional closure* (the EEC): this is one of the central concepts of Sonata Theory and one that is dealt with at length in other chapters.[6] Producing the EEC is the generically assigned task of the S-idea(s). The large dotted-line arrow in figure 2.1a suggests a broadly vectored trajectory from the start of the exposition to the EEC; the smaller dotted-line arrow below it suggests a subordinate trajectory from the beginning of S to its own point of PAC-closure at the EEC. In performing or listening to any sonata-form exposition one should sense the broad drive of these generic vectors. Whenever one hears the onset of S-space within any exposition, one should listen with an alert sense of anticipation for any subsequent PAC—how it might be approached, secured, delayed, thwarted, or deferred. One should experience any sonata form with a strongly "directed" preparatory set, pressing forward conceptually and anticipating genre-defining events-to-come.

Following the EEC one or more additional cadences (PACs) may follow within the *closing zone* or *closing space* (C). (Not all expositions contain C-modules; it is possible for the S-concluding EEC to be delayed until the end of the exposition, in which case there is no closing zone.) Whether or not C-modules are present, the final cadence of the exposition will generally be a perfect authentic cadence in the secondary key (again, V:PAC, III:PAC, or v:PAC). This final cadence might not occur directly at the double bar. Frequently the final cadence is followed by a C-module that prolongs the newly reinforced tonality by means of a pedal-point or some other device. Additionally, the final cadence is sometimes followed by a reactivation of V in preparation for a repeat of the entire exposition: if so, this reactivating passage is the *retransition* (RT).

Development

This action-space renders the established tonal tension more fluid and complex. While the exposition had split its tonal assertions into two broad blocks or contrasting planes (I and V in major-mode sonatas), the development typically initiates more active, restless, or frequent tonal shifts—a sense of comparative tonal instability. Here one gets the impression of a series of changing, coloristic moods or tonal adventures,

6. For the moment, we might emphasize that the first *satisfactory* PAC in the new key is often but not always the *first* PAC in that key. A first PAC, for instance, might be followed by a thematic repetition of all or part of the S-idea that we have just heard—which would automatically defer the EEC to the next satisfactory PAC further ahead. Additionally, there are other ways of deferring the sense of a clear EEC (ch. 8). The clearest way of suggesting all of this in brief is to define the EEC as the first new-key PAC that *proceeds onward to differing or contrasting material*—or, of course, that closes the exposition itself, if there are no closing modules that follow that PAC.

often led (in major-mode works) through the submediant key, vi, or other minor-mode keys with shadowed, melancholy, or anxious connotations. Any authentic-cadence attainment in a non-tonic key is to be understood as an important developmental event—a cadential ratification of an attained tonal station. (A vi: PAC is especially common in major-mode sonatas.) Ultimately, the standard development culminates on an active dominant (V_A, meaning "a V that is an active chord, not a key"). At this point the dominant from the end of the (major-mode) exposition is usually recaptured, detonicized, and reactivated.

This last point needs underscoring. In the development the final cadence is usually a half-cadence in the tonic (I:HC), although a cadence in a related minor key, normally followed by a brief reactivation of V, is also a possibility. In addition, a I:HC is frequently followed by a prolongation of dominant harmony, a "dominant-lock" or "dominant preparation." The typical I:HC conclusion of the development—just before the onset of the recapitulation—brings us to a harmonic *interruption*. (This crucial interruption is a defining feature of the Schenkerian conception of sonata form.) The V_A at the end of the development is not resolved to the I that usually begins the recapitulation. Rather, the phrase—and the development section as a whole—is normally "interrupted" on V_A (notwithstanding any foregrounded or local, connective "fill" that might bridge the end of the development to the recapitulation), and the next cycle of events is newly launched with the opening of the recapitulation. True, this more fundamental interruption on the dominant may sometimes be masked on the foreground with an *apparent* V–I cadence (with the I triggering the recapitulation). But the more fundamental or background concept is that of harmonic interruption on V_A. (Those unfamiliar with the Schenkerian, linear-contrapuntal view of things might notice that this interruption divides the entire sonata form at the end of the development. This contrasts with the eighteenth-century "binary" division of sonata form at the end of the exposition.)

In terms of their *rhetorical* strategies, developments may or may not be fully or partially *rotational* (that is, guided in large part by the ordered thematic pattern established in the exposition). Developments often refer back to (or take up as topics) one or more of the ideas from the exposition, most commonly selected, as it happens, from Rotation 1's first half (P and TR). More often than not, the modules taken up and worked through in the development are presented in the order that they had originally appeared in the exposition (even though several expositional modules are normally left out entirely). Thus the modular succession encountered in the development—not only the expositional events referred to, but also the possibility of an episode or largely new theme—is never to be considered arbitrary. On the contrary, even within this more unpredictable, developmental texture the thematic choice and arrangement is of paramount importance and derives its significance through a comparison with what had happened in the exposition. The development is variable in length, although in the period 1760–90 one would normally expect it to occupy a smaller space than that established by the exposition. Longer, more elaborate developments in the 1780s, 1790s, and later decades are monumentalized statements that invite special attention.

Recapitulation

This action-space resolves the tonal tension originally generated in the exposition by rebeginning on the tonic (with the initial theme in the most common Sonata Types, 1, 3, 4, and 5) and usually by restating all of the non-tonic modules from part 2 of the exposition (S and C material) in the tonic key. For this reason—its largely referential retracing of the rhetorical materials laid out in the exposition (Rotation 1)—we also call the recapitulation the *recapitulatory rotation*. (Exceptions and reorderings of thematic material may be found in some sonatas.) Because of its function in bringing tonal closure to the entire form, we refer to the S/C complex in the recapitulation as the *tonal resolution*. Its shape and manner of unfolding had been established by the exposition's structure of promise. Correspondingly, we consider the recapitulation to articulate a *structure of accomplishment*. Minor-mode sonatas

that had sounded S and C in the major mediant (III) in the exposition have the additional option of sounding them in either the major or minor mode in the recapitulation.

The recapitulation's S, launching the tonal resolution following a recapitulatory MC, leads to the production of a satisfactory I:PAC that goes on to differing, non-S material. This is the moment of *essential structural closure* (the ESC), most often a point parallel to the exposition's EEC. The ESC represents the tonal goal of the entire sonata form, the tonal and cadential point toward which the trajectory of the whole movement had been driving: this is suggested by the longest dotted-line arrow in figure 2–1b. From the perspective of Sonata Theory, it is only here where the movement's tonic is fully called forth, stabilized as a reality as opposed to a mere potential. As in the exposition, C-material will follow, now in the tonic. The recapitulation's final cadence is generally a I:PAC (or, in minor, sometimes a i:PAC), although this too may be followed by a prolongation of tonic harmony or by a transition leading either back to a repeat (of the entire development and recapitulation) or forward into the coda.

A coda (outside of sonata space) may or may not follow the recapitulation. More information about codas, along with a discussion of the other optional or parageneric feature of some sonatas, the introduction, may be found in chapter 13.

Repetition Schemes

Within eighteenth-century sonatas and symphonies one may find both parts repeated (||: exposition :||: development–recapitulation :||). This is the most formal and earliest norm. Many late-century first movements, especially those after about 1760, repeat only the first part (the exposition), although in works prior to 1790 one need not be surprised to see the second part also repeated. After that date, repeating the second part is an uncommon gesture that invites analytical interpretation. It is also possible to find both parts unrepeated. This occurs in lighter works, in some midcentury symphonies (some Stamitz symphonies from the 1750s; some early Mozart symphonies; and so on) and in some slow movements (especially those in the format of the less expansive, Type 1 sonata, lacking a development). The nonrepeated exposition is also a generic feature of the overture or *sinfonia*. (In other words, expositional repeats will not appear in either operatic or concert overtures; this is also true of the overture's mid-nineteenth-century offspring, the symphonic poem).[7] In this aspect the lighter overture is to be distinguished generically from the more formal first movement of a sonata or grand symphony, which at least had available the common option of expositional repetition. Nonrepeated expositions within first movements do sometimes occur in more broadly scaled and ambitious works after 1780, but when they do—as in Mozart's Symphony No. 35 in D, K. 385, "Haffner,"[8] or in Beethoven's Violin Sonata in C Minor, op. 30 no. 2, his Piano Sonata in F Minor, op. 57, "Appassionata," and his String Quartet in F, op. 59 no. 1—they are exceptional and need to be considered as consciously expressive choices.[9]

One curious (and rare) possibility is that of literally writing out an expositional repeat,

7. Thus the rule. Exceptions are extremely rare and disconcertingly puzzling, such as the repeat of the exposition in young Mozart's Overture to *Apollo et Hyacinthus*, K. 38 (1767), labeled as the "Prologus/Intrada" to the opera. This piece is a Type 2 sonata (Chapter 17) whose first rotation (exposition) is provided with a repeat sign. Much later, the odd "expositional" (?) written-out and slightly varied repetition in Berlioz's Overture, *Le carnaval romain* is also curious, suggesting that the form of this unusual piece is more purely rotational (or perhaps instrumental-strophic with *fortissimo* refrain) than a sonata per se, although it is also manifestly in dialogue with certain sonata norms.

8. Other examples within Mozart's major works include the first movements of his Symphonies Nos. 31 in D, K. 297, "Paris," and 34 in C, K. 338, along with those of the Serenades in D, K. 320, "Posthorn," and in E-flat, K. 375. Such examples—perhaps related to earlier or existing concepts of repeat-convention options in overture-symphonies, in smaller-scale symphonies, or in some serenades—require individual attention. Within the larger symphony it may be that during the 1770s (though not, it seems, in the 1780s) Mozart was exploring the possibility of the omission of the expositional repeat as a lower-level default.

9. The solution of Beethoven's op. 59 no. 1/i, which initially suggests an expositional repeat only to abort it almost immediately in favor of development, is anticipated in the first movements of Mozart's Serenade in E-flat for Eight Winds, K. 375, and Haydn's Piano Sonata in D, Hob. XVI:51.

normally including variants the second time around. This occurred most famously in C. P. E. Bach's unusual set of six keyboard *Sonaten mit veränderten Reprisen,* H. 136–39, 126, 140 (W. 50/1–6, Sonatas with Varied Repetitions), composed in 1758–59 and published in Berlin the following year. In Haydn's works the procedure surfaces only (and wisely, in Tovey's view) in a few "purely lyric slow movements," such as the Adagios of the Quartet in C, op. 33 no. 3, "Bird," and the Symphony No. 102 in B-flat.[10] (Both slow movements are in F major; in the quartet the Adagio is the third movement; in the symphony it is the second.)

What are the purposes of large-scale repeats within sonata form? Central to the concept of the grand sonata or symphony is a system of schematic repeat-conventions, balances, symmetries and proportions that call attention to and help to define the genre. The emphatically architectural construction calls attention to the genre's ordered formality—and in the case of the grand symphony, also to its grandeur and public splendor. Repeats were an important feature of a sumptuous, high-prestige display of grand architecture, one to which large-scale repetitions were essential —especially that of the expositional repeat in the first movement. The stylized form thus celebrated the "Enlightenment" (or "modern") culture that makes such an impressive, moving, or powerful art possible. One of the structure's implications would have been that this culture had devised a rational, balanced means to shape and contain the fluid, raw, elemental power of music. By extension, the process probably also represented the controlling or harnessing of those impulsive, instinctive, libidinal, or "uncivilized" elements within ourselves. Control, balance, generic identification, and formal architectural splendor: these would appear to be the central reasons why literal repetition played such a prominent role in the style.

Consequently, repeat signs should not be taken for granted, passed over lightly in analysis, or omitted in performance. Repeat signs are never insignificant.[11] Block-repetitions are an integral component of the style, and composers can work with this defining convention in a variety of ways. When previously obligatory (or exceptionally strong first-level default) expositional repeats began gradually to disappear—especially in the early nineteenth century, with certain works of Beethoven (op. 30 no. 2, op. 57, op. 59 no. 1, and so on, and later with Mendelssohn, Schumann, and others)—the genre itself was undergoing a major rethinking.[12] The familiar, current views—Schenkerian and otherwise—that propose that some repeats are structurally insignificant while others are more important (because of the unfolding of certain structural tones or other significant events, perhaps under a first-ending sign) miss the larger point of repeat signs as generic identifiers.[13] Even when the structural-tone aspects might be convincing (but, perhaps paradoxically, only as local details), the gist of these claims seems to be based on later-nineteenth-century premises, which came to look on all unaltered repetition as an aesthetic error. Such a conviction also came to affect performance in the omission of repeats or in the insistence on an altered interpretation in the repeat. It may be, though, that saying the same thing twice was what the composer had in mind.

It is easy to object to our general argument here. One could strive to minimize the impor-

10. Tovey, "Sonata Forms," *Musical Articles*, p. 214: "Haydn saw that the only place for C. P. E. Bach's device was in purely lyric slow movements. Even there he never had the patience to plod and pose (as C. P. E. Bach did to the bitter end) through a repetition [recapitulation] of both parts. When his second part comes to recapitulate the second group it combines both versions."

11. For the quintessential statement of that which the present argument opposes, see Douglass M. Green, *Form in Tonal Music: An Introduction to Analysis*, 2nd ed. (New York: Holt, Rinehart, and Winston, 1979), p. 82: "HISTORICAL NOTE [*sic:*] Ordinarily the repetition of a part is of little significance in formal analysis."

12. Curiously, in 1826 Reicha suggested—in passing and without explanation (*Traité de haute composition musicale*, p. 300)—that finales may lack an explicit repeat: "When the first part is not repeated, as in overtures and finales . . ." ("Quand la première partie n'a pas de reprise, comme dans les ouvertures ou dans les finales . . ."). It may be that Reicha had sonata-rondos, Type 4 sonatas, in mind (ch. 18).

13. Cf., e.g., Jonathan Dunsby, "The Formal Repeat," *Journal of the Royal Musical Association* 112/2 (1987), 196–207.

tance of the usual repetition schemes by an appeal to history: deriving them step-by-step from the earlier binary forms, then asserting that the persistent lingering of the repeat conventions into the 1780–1820 period of the grand symphony was an outdated survival, vestigial, unnecessary to the perception of the genre. The larger question, though, is why the convention remained available into the later phases of 1780–1820 period and beyond (particularly after Beethoven's occasional removals of the expositional repeat had occurred). The expositional repeat must have persisted, however sporadically, because it was not merely vestigial. It continued to be genre-defining, a sign of special grandeur and formality—with an ear attuned also to the grand tradition and historical lineage that had led to the mid- and later-nineteenth-century sonata and symphony.

Of the two standard large-scale repeats, the second, longer one (development-recapitulation) was the one more vulnerable to suppression. This second repetition was increasingly reduced to the status of an easily discardable option in the 1780–1800 period.[14] In some cases, concerns of absolute length or a sense of redundancy in closing particularly dramatic sonatas twice might have overridden the genre-defining principle of long-range architectural repetition and balance, at least with regard to this development–recapitulation section. Perhaps the logic of the situation suggested that the obligatory repeat of section 1 alone (the expositional repeat) was to be viewed as sufficient as a genre definer.

However we decide this matter, we should note three things. First, the issue of notationally indicating a repeat of section 2 was still part of the historical concept of "grand binary" form (within a symphonic first movement) around 1800, even when that repeat was notationally elided. Its conceptual presence remained there, counterpointed against the given, simpler structure. It persisted as historical-generic memory, even when it was not made physically present on the acoustic surface of the music. Second, any retention of the second repeat toward the end of the eighteenth century should be regarded as expressively significant, especially since its strongest composers—Haydn and Mozart—were apparently coming to believe that repeat 2 was not as obligatory as that of repeat 1. When the repeat was called for, it must have been placed there for a reason, as in the slow movement and finale of Mozart's Symphony No. 41 in C, K. 551 ("Jupiter"), where formal processes and monumentalized grandeur are principal topics throughout the whole work. Third, given a nineteenth-century work lacking an indication of that second block-repetition, any reworked referencing back to this increasingly atavistic repeat 2 within a longer, discursive coda, as in the first movement of Beethoven's Symphony No. 3 in E-flat, op. 55, "Eroica," should be viewed as such, not as an innovative addition or accretion to a previously postulated, differing symphonic practice.

14. See, e.g, Michael Broyles, "Organic Form and the Binary Repeat," *The Musical Quarterly* 66 (1980), 339–60.

CHAPTER THREE

The Medial Caesura and the Two-Part Exposition

An exposition has both a tonal and a rhetorical function. Its basic tonal plot—moving from an initial tonic to a secondary key, then securing that new key with one or more cadences—constitutes the exposition's *tonal form*. This tonal form, generally the same in all sonatas, is worked out in different sonatas in individualized ways, according to localized rhetorical plots. Tonal form is to be distinguished from *rhetorical form*, which includes personalized factors of design and *ad hoc* expression: modular and textural layout, selection and arrangement of musical topics, varieties of structural punctuation, and so on. The compositional ordering of these processes produces a distinct, singular musical shape. This layout serves as the *referential rotation* (or *expositional rotation*) that also guides our understanding of the ordering of modular events in the subsequent action-spaces of the sonata—development, recapitulation, and coda. An exposition may be disposed in either of two rhetorical formats: the *two-part exposition* (containing a medial caesura) or the *continuous exposition* (lacking a successfully articulated medial caesura). In this chapter we are concerned with the former.

The Two-Part Exposition

This is the format most frequently employed by most composers of the second half of the eighteenth century. Hence when one confronts any sonata form from this period, the most reasonable initial expectation would be that one is about to encounter a two-part exposition. (As will be seen in chapter 4, a continuous exposition often plays upon, then overrides, this expectation.) The cardinal feature identifying this exposition type is the presence of a sufficiently deployed medial caesura and (often contrasting) second theme.

Part 1 comprises the establishment of the tonic and the energized drive to the medial caesura. It contains two action-spaces, the *primary-theme zone (P)* and the *transitional zone (TR)*, and culminates in the *medial caesura (MC)*, which we indicate by an apostrophe (’). Part 2 comprises the post-MC material and lasts until the end of the exposition. This section is concerned with the cadential affirmation of the new key (V in major-key sonatas, III or v in minor-key ones). Part 2 subdivides into the *secondary-theme zone (S)*—which normally concludes with the sounding of the first satisfactory perfect authentic cadence (PAC) in the new key that proceeds onward to differing material, the event that we

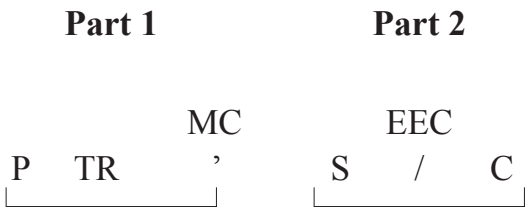


FIGURE 3.1 The Two-Part Exposition

call the moment of *essential expositional closure* (EEC), indicated here by a slash (/)—and the *closing zone (C)*. The two-part exposition may be represented as in figure 3.1.

When beginning the analysis of any exposition, we recommend that the first task be to locate and identify the treatment of the MC—to determine, first, if one exists at all and, if so, to investigate what kind it is, where it falls within the exposition, what complications might surround it, and whether the moment identified actually leads to an acceptable secondary theme (S). The second task should be to examine the strategy surrounding the EEC. Productive analyses often start in the middle of the exposition and work outward to the beginning and the end.

The Medial Caesura (MC): Definitions and Overview

The *medial caesura* is the brief, rhetorically reinforced break or gap that serves to divide an exposition into two parts, tonic and dominant (or tonic and mediant in most minor-key sonatas).[1] (A touchstone occurrence of this familiar break may be consulted in example 3.1 below: the first movement of Mozart's Piano Sonata in D, K. 284, occupying all of m. 21, with the literal gap on beat 4.) In rapid-tempo compositions a medial caesura is usually built around a strong half cadence that has been rhythmically, harmonically, or texturally reinforced. The half cadence proper—the moment of cadential arrival on an active dominant—often occurs before the MC itself. Very commonly, this active V (V_A) is then prolonged, kept alive, for several more measures as an actual or implied dominant pedal-point, a *dominant-lock*, driving aggressively toward the MC articulation. Thus while the moment of the MC proper—the articulation of the gap—is frequently not literally identical with the moment of the half-cadence arrival (the HC, which could have happened several bars earlier), the larger drive to and execution of most MCs are nonetheless "built around" a half-cadence effect or "dominant-arrival effect" in either the tonic or the dominant key.[2]

Viewed broadly, the entire process from the half-cadence arrival proper through the literal execution of the terminal MC-break, which might occur several measures further ahead, expresses a purposefully activated and prolonged half-cadence effect. In referring to medial caesuras as often being "built around" half cadences, we of course distinguish between the point of initial half-cadence arrival and the MC moment itself, in those cases where these two events differ. Nevertheless we also use the shorthand symbols I:HC MC (a medial caesura that often terminates the sustaining of an active V in the tonic) or V:HC MC (one that often terminates the sustaining of an active V in the dominant) to suggest this whole complex of musical activity, one in which the literal MC moment is to be interpreted referentially to any preceding moment of half-cadence arrival.

---

1. Much of what follows is adapted and updated from Hepokoski and Darcy, "The Medial Caesura and Its Role in the Eighteenth-Century Sonata Exposition," *Music Theory Spectrum* 19 (1997), 115–54. Some of the adaptations seek to clarify and make more precise issues raised regarding that article by William E. Caplin in "The Classical Cadence: Conceptions and Misconceptions," *Journal of the American Musicological Society* 57 (2004), 51–117. While we do adopt some of Caplin's cadential terminology—identifying a moment of "cadential arrival," for instance—our view of the activity surrounding the MC-event differs from his in several respects, as will emerge.

2. Defining precisely what consitutes a half cadence is no easy matter. For one version of the concept of dominant arrival, as opposed to a half cadence proper, see Caplin, *Classical Form*, pp. 79–81. Cf. nn. 6, 11, and 14 below, on the claim that the dominant-lock is best regarded as "postcadential."

The medial caesura has two functions: it marks the end of the first part of the exposition (hence our adjective "medial"), and it is simultaneously the highlighted gesture that makes available the second part. The MC is the device that forcibly opens up S-space and defines the exposition type. Somewhat whimsically, it may be thought of as metaphorically analogous to the moment of the opening of elevator doors onto a higher floor—making S-space possible or opening to the second part of the exposition. The medial caesura provides a firmly established platform from which the secondary theme, launching part 2, may emerge. In order for the MC to do its job most effectively within rapid-tempo compositions, energy must be applied. This energy is furnished by TR, the transitional zone. As a rule of thumb, once TR has begun, the *forte* energy should be kept constant or on the increase all the way to the medial caesura proper. Any flagging of energy or vigor within TR—any *diminuendo* or faltering drop to *piano*—is countergeneric and constitutes an event that invites interpretation. It may suggest the production of something unusual: a medial caesura deformation or the presence of a troubled expressive problem being unfolded in the musical narrative.

The Medial Caesura: Harmonic Defaults

As indicated above, the MC is most commonly the final gesture—the "break" or "gap" at the end—of a more complex musical passage constructed around and often sustaining a half cadence (HC) or dominant arrival, in either the tonic key (I:HC) or the dominant key (V:HC). As will become clearer as we proceed, I:HC MCs are generally more appropriate for shorter, lighter pieces. On the other hand, V:HC MCs tend to be more frequent within ambitious works of moderate length and larger, especially toward the end of the eighteenth century. As such—and even though one can come across exceptions to the proposed idea—encountering either the one MC or the other can be a signal both of the level of complexity at hand and of the probable proportions of what is to follow. A I:HC MC helps to predict a work on a relatively modest scale; a V:HC MC—suggesting a more harmonically complex option—proposes an exposition with "grander" proportions. (Such considerations are related to the concept of the "deployment sequence of medial caesura options," discussed separately below: a composer sometimes seems to pass up a I:HC MC possibility in order to pursue a V:HC MC down the road.)

The later eighteenth century saw a general increase in the expansiveness and ambition of individual movements. As a result, in most major-mode cases the MC is constructed around a half cadence or active-dominant arrival in the dominant key: the familiar V:HC option. Because of this statistical frequency, we refer to it as a *first-level default* for this expositional moment when we are dealing with works of at least moderate length.[3] (For more unassuming pieces, on the other hand, one might argue that the I:HC MC could be the first-level default or "most obvious" choice.) In the case of the V:HC MC the transition will have modulated from the original tonic to the dominant. In many transitions the preparation for the major-mode S (in V) is accomplished through a darkened or stressful pathway in the parallel minor of that new-tonic-to-come. In such situations, therefore, the TR-drive to the V:HC MC is produced with a concomitant shift to the minor mode. This means that the moment of the MC is locally sounded as the terminal gesture of a prolonged half-cadence-effect in the minor dominant, *v:HC,* whereupon S follows, more brightly, in the major mode. (A particularly charged instance of this dramatic *chiaroscuro* may be found in the first movement of Beethoven's Symphony No. 2 in D, op. 36, with mm. 61–71 delivered aggressively on A minor [MC at m. 71] and S emerging in A major at m. 73, after

3. An array of statistical evidence regarding the frequency of what we call V:HC and I:HC medial caesuras in Haydn, Mozart, and others has been compiled in Robert S. Winter, "The Bifocal Close and the Evolution of the Viennese Classical Style," *Journal of the American Musicological Society* 42 (1989), 275–337. (We, however, do not find the term "bifocal close" for the I:HC MC to be helpful.)

two bars of major-inflected caesura-fill.) Such occurrences participate in a dialogue with the normative, major-mode MC expectation. The v:HC MC option is a commonly elected negative overlay onto the conceptual first-level default, V:HC. We still have a first-level default MC, but one subjected to the additional surface feature of temporary minor-mode mixture—a momentary "lights-out" feature. (Chapter 14 considers the expressive implications of such a mixture.)

Additionally, within the first-level default in either major-mode or minor-mode expositions one occasionally finds a seventh included in the V chord at the MC point. This seventh is best regarded as a passing tone. In the first movement of Haydn's Symphony No. 100 in G ("Military"), for example, the structural V/V, the half cadence proper, is articulated at the downbeat of m. 62, after which it is prolonged. During this prolongation the seventh is added (entering first in mm. 64–65, though most prominently in mm. 69–73), suggesting a V^{8-7} figure. The seventh ($\hat{4}$ of the new key) resolves to an inner-voice $\hat{3}$ at the onset of S (m. 75). This addition of the seventh during the drive to the MC proper is not uncommon but is normally limited to the first-level default (V:HC MC). If the seventh were added to the dominant of a second-level default (I:HC), the tendency of the resultant V^7 to resolve to the tonic would preclude the requisite tonal shift to the key of the dominant. Nevertheless, this can occur at the MC point of the initial tutti rotation of a Type 5 sonata (concerto movement), where a modulation is not required, as in Mozart's Piano Concerto No. 21 in C, K. 467, mm. 20–26.

On rare occasions one encounters the substitution of an inversion for the V or V^7 chord at the MC point. Regardless of whether the dominant has previously appeared in root position, this situation should be understood as a *medial-caesura deformation*, which might well impact on the subsequent S. In Beethoven's C-minor *Coriolan* Overture a strenuous TR manages to lock onto V^6 of E-flat (here entering as E-flat minor) in m. 46. (Obviously, this cannot be construed as a half cadence proper.) After four measures of convulsive upheavals around that chord, a V^6/iii MC-effect occurs in m. 50. Two bars of descending caesura-fill in the first violins lead to the E-flat-major start of S in m. 52. In this case the rhetorical effect produced is that of the half-cadence MC—the passage is clearly in dialogue with that norm—but the music is dramatically staged as being "unable" to produce the more normative HC at this point. Somewhat similarly, in Mozart's C-major Overture to *La clemenza di Tito*, K. 621, a suddenly introduced V^6_5 /V (m. 28) seems to startle the music into nothing less than a fermata-stop (m. 29) that serves as the exaggerated GP-gap of a V^6_5/V MC deformation. S begins in G major in m. 30.

At least in works of substantial length the second most common major-mode option, the *second-level default*, is to build an MC around a half cadence or dominant arrival in the original tonic, a *I:HC*. (As mentioned above, in shorter works, a case might be made that the I:HC MC is somewhat more appropriate: it might be regarded as a first-level default in certain situations. For the present, the discussion is framed around works with grander proportions.) In this second-level default, I:HC, the transition will not have modulated: it will have begun and ended in the tonic, and it will be up to the ensuing S-space to establish the new key, usually by beginning directly in it. Because second-level MC choices are not infrequent, TR-space cannot be defined in terms of modulation. Once again, first- and second-level MC defaults are not expressively equivalent. The first, V:HC MC, is a more decisive gesture: it announces the intention to open part 2 more solidly, with its new key already in hand. The second, I:HC, is weaker, usually occurs early on, predicts a briefer or less ambitious sonata, and sometimes purposefully generates problems in what follows.

The minor-mode, derivative analogues to the above are III:HC or v:HC as strong first-level defaults, depending on the key to which one is modulating (moving to the minor dominant occurs much less frequently), and i:HC as the second-level default. While examples of the former are frequent—an MC built around a half cadence in the new key is the most common choice—examples of the latter are relatively rare. The reason why is obvious. While major-mode statements of the I:HC MC (sec-

ond-level default) may easily become the tonic chord of the new key (V) and the S-to-come, this is not the case in minor-mode works. In other words, i:HC MC (say, a G-major chord sounded as an active V in C minor) is obliged to yield at once—as a *quasi non sequitur*—to the tonic of the mediant major with the onset of S (the key of E-flat in a C-minor exposition, whose appearance also produces a cross-relation between the original dominant chord's B natural and the new tonic's B flat). When this does occur, the effect can be striking: a sudden pull out of the ominous tonic minor into the brighter, more "hopeful" mediant major.[4] The touchstone examples may be found in the Prestissimo finale of Beethoven's Piano Sonata in C Minor, op. 10 no. 1 (i:HC MC at m. 16, stalled with a quizzical fermata, followed directly by the forthright S in E-flat at m. 17) and in both the opening ritornello and solo exposition of Mozart's Piano Concerto No. 20 in D Minor, K. 466 (a literal i:HC at m. 28, proceeding into a dominant-lock and i:HC MC at m. 32 and a sudden shift to F major, III, for the opening of the ritornello's secondary-theme space at m. 33; see also the solo exposition, with its i:HC MC at m. 114). From time to time the production of a i:HC MC can be presented as a compositional problem that needs immediate emendation. In the first movement of Beethoven's Quartet in C Minor, op. 18 no. 4, a i:HC MC-effect in m. 25 leads not to S proper but to a classic situation of "medial caesura declined" and the initiation of one type of trimodular block (TMB)—both of which strategies are discussed separately below. Decades later, Mendelssohn would provide the B-minor *Hebrides* Overture with a i:HC MC (m. 43; notice also the earlier i:HC MC-effect in m. 39, which is returned to and recaptured in m. 43) and assign the modulation to III to the "poetic" caesura-fill bridging the MC to the onset of the *cantabile* S in D major (m. 47).

Most sonata forms display either a first- or second-level default MC, one built around a half cadence that may or may not have been prolonged by means of a dominant-lock. Much less frequently, one may find an MC-function produced by a perfect authentic cadence in the new key (PAC). In major-mode sonatas this *third-level default* is V:PAC MC, which occurs, for instance, in the first movement of Mozart's Quartets in D, K. 155, m. 28; in E-flat, K. 428, m. 40; and in B-flat, K. 589, m. 45. This procedure is also found with some frequency in earlier and briefer works. (The V:PAC MC in the first movement of Haydn's Symphony No. 10 in D, m. 23—one of several examples in early Haydn—could hardly be clearer.) The minor-mode-sonata analogue is normally III:PAC, as in (though with subsequent complications) the first movement of Mozart's Quartet in D Minor, K. 421, m. 24 (see example 4.3 and the accompanying discussion in chapter 4). Alternatively, in works that shift to the minor dominant key, one might find a v:PAC MC.[5]

PAC MCs are stronger tonal and rhetorical gestures than are HC MCs. Because they are heard as signs of closure, not of expectancy, and because they sound the same perfect authentic cadence that will define the EEC concluding the secondary theme, they present problems of understanding. From time to time they emerge after a composer has already dallied with the V:HC option—perhaps already producing an HC arrival in that key, or very nearly so. When this happens, it is as though the music at first

4. In effect the juxtaposition is produced by a chromatic 5–6 shift, in which both the third and the fifth of the active V (the G-major chord within a C-minor exposition) are altered. Recent neo-Riemannian theory might also describe this as a PL shift: a simultaneous application of a color-shift to the parallel (P) mode (a G-major chord, V of C minor, thus inflects to G minor) followed by a "leading-tone exchange" or *Leittonwechsel* (L) (the resultant G-minor sonority inflects its fifth, D, to E-flat, thus producing an E-flat chord). This familiar juxtaposition is discussed in somewhat more detail in ch. 10 in conjunction with the final active-dominant chord of the recapitulation, where the more typical V_A sonority is sometimes replaced by V/vi (V of A minor within a C-major movement, for example, giving way almost immediately to the C-major recapitulation).

5. An extremely rare—and clearly deformational—alternative is found in Beethoven's D-minor Largo e mesto slow movement of the Piano Sonata in D, op. 10 no. 3, which features a VII:PAC MC (a PAC on V of III, C major, in m. 17)—as if seeking to "close" early, albeit in the "wrong" major mode. The S and C that follow are in A minor (v).

"decides" to drive toward the normative V:HC MC, only suddenly to "change its mind," abandoning the normative implications of the dominant-lock (if that lock had indeed been initiated), and pushing instead, impulsively, toward the stronger PAC in the new key. Thus a V:PAC MC is sometimes produced in a context that has suggested, then overridden, a more normative half-cadence-effect MC option.[6] (In K. 155/i, for instance, the half cadence and dominant-lock, V:HC, occurs at m. 20. The lock proper sustains the V_A through mm. 20–23 but is abandoned at m. 24 in order to plunge into the emphatic V:PAC MC at m. 28. The subsequent S theme, m. 29, as it happens, appears to register its surprise by beginning off-tonic, on the supertonic chord.) In some of these instances V:PAC MCs are elaborate, more decisive versions of caesura-fill of the $\hat{5}-\hat{1}$ -descent type, and distinguishing between the two can be difficult or very much a matter of individual interpretation. (This last feature is discussed in more detail below, in the discussion of caesura-fill.)

Such observations lead to larger speculations: the very concept of a V:PAC MC is potentially problematic. What leads us to think that what we call a V:PAC MC is not the EEC? (Such an early EEC would define that exposition as continuous, not as two-part, since there would have been no prior MC.) One might reason that if the generic goal of an exposition is to produce a satisfactory PAC in the secondary key, any such V:PAC-effect at this point might initially lead us to suspect that that aim has been achieved. One soon learns that it is part of the expressive character of a local V:PAC MC to threaten to preempt the sense of tonal closure that we associate with the EEC—the V:PAC that must be re-produced down the road at the end of S. But how certain can we be that such a V:PAC should not be taken for the EEC? This decision is a crucial one. It concerns the structural importance of that first V:PAC at what might well be the MC: is it a secondary, local effect (perhaps describable as the strongest possible "tonicization" of V of I at the MC point)? Or is it to be taken as a decisive structural event within the genre—nothing less than the EEC? Or can there even be other options for interpretation?

Deciding this matter in individual cases depends on three factors, each of which involves matters of interpretation and experience. The first is the question of how far into the exposition the V:PAC is sounded. Once we have proceeded past about 65 or 70 percent through—in other words, once we have experienced a proportionally overlong transition—V:PACs become less convincing as MCs, since a medial caesura normally occurs earlier in the exposition. The later the V:PAC is produced, the more likely it will be taken as the EEC. (Put another way, normally the only way that a composer can have a V:PAC serve as an MC is to expand the proportions of what follows to the point where it can be regarded as a convincing part 2: S and C. For this reason the V:PAC MC option typically suggests an exposition and subsequent sonata of notable proportions.) The second factor concerns how the V:PAC was prepared and produced (idiosyncrasies in the preceding TR). The third is the character of the module that follows. If it is a clear, contrasting theme, is it S-like or

6. In response to one case of this as illustrated in our article, "The Medial Caesura," pp. 129–30 (Beethoven, Piano Trio in G, op. 1 no. 2, first movement, mm. 98–99), Caplin ("The Classical Cadence," pp. 108–12) insisted that the PAC-effect at m. 99, while displaying "cadential *content*" (the two-chord combination V^7–I), could not be considered a PAC proper because it lacked "cadential *function*," at least according to his "highly constrained" (p. 56), much-restricted definition of that function. In Caplin's view, once an HC-arrival has been attained in TR, all that follows in the dominant-lock must *ipso facto* be considered "postcadential" and by that definition incapable of producing a PAC at its end. That argument, however, fails to consider the possibility that a dominant-lock might be abandoned *en route*—in other words, that it might be staged as "changing its mind"—in order to proceed to a PAC. Such a procedure would "unfreeze" the locked dominant (still an active dominant, a V_A, after all) and treat it as more of a "normal" V_A that can proceed onward toward resolution. In any event, we agree that Beethoven's op. 1 no. 2/i provides a problematic case along these lines, since the lurch to the V:PAC happens so rapidly and so closely resembles caesura-fill of the $\hat{5}-\hat{1}$ -descent type. The following example mentioned in the text, Mozart's K. 155/i, provides a clearer illustration of the process. Some of these issues are revisited in n. 14. Cf. also n. 2 and nn. 11 and 14.

C-like? Deciding this is not always easy. This general situation arises with some frequency in Haydn, who had a fondness for planting a decisive V:PAC in the 55–70 percent range of the exposition.[7] Alternatively, what follows the first V:PAC might be less of a "theme" than one or more short modules that recapture or restate that mid-expositional cadence. When that occurs, we are dealing with what we regard as the second type of continuous exposition, a possibility dealt with in chapter 4.

The even rarer option, a I:PAC MC, may be considered a *fourth-level default*. In eighteenth-century works a I:PAC or IAC-substitute (imperfect authentic cadence) leading to an obvious S in the new key may occasionally be found in light, small-scale works, in some telescoped or abbreviated expositions, and in some slow movements. Generally the PAC or IAC closes off a brief, straightforward P, and the resulting impression is that of omitting the TR-zone altogether. Because of the effective ellipsis of TR, the I:PAC or IAC at the end of P is asked to do double duty as the rhetorical MC. This occurs in the first movement of Mozart's Symphony "No. 7" in D, K. 45, m. 16 (I:IAC MC, with S in V at m. 17) and his String Quartet in A, K. 169, m. 11 (I:PAC MC, with S in V at m. 12).

In larger, more ambitious pieces the extremely infrequent I:PAC or IAC MC can carry a different implication. Here, following P, one enters what seems rhetorically to begin as a normative TR. That TR, however, proves unable (or unwilling) to produce any of the three more standard MC defaults: V:HC, I:HC, or V:PAC. In some expositions it may "try" to produce one of those—or to initiate a motion in one of those directions—before being drawn back to the original tonic. In other cases it may simply bask in an ultra-stable tonic without any gesture toward a typical MC. In either situation one confronts a "failed" (or gesturally weak? or obstinate?) TR that, still in the grip of the grounding tonal principle of the P-zone, dwells on an unusually static tonic. This emphasis, in turn, demands analytical and hermeneutic interpretation.

The classic example occurs in the first movement of Mozart's String Quintet in G Minor, K. 516. In this extraordinary exposition the negative pull of G minor is apparently so strong that TR (beginning in m. 9 as a TR of the dissolving-consequent type) finds itself unable to escape its control. The result is one of the bleakest MCs in the repertory, the i:PAC at m. 29. The preceding, *forte* i:PAC at m. 24, Neapolitan-enhanced (m. 23) and brusquely closing the door on the fatalistic G minor, foreshadows this MC-effect. What intervenes in mm. 25–29 is a "timid," failed attempt to wrest free of the clutches of G minor through a momentary glance at VI. Being drawn back once again to G minor and to the i:PAC in mm. 27–29 is chilling—a second confirmation of the countergeneric inability to escape from the gravitational negativity of the tonic. The S that follows in m. 30 (the rhetorical signals make it clear that this is S) begins in the same, inescapable G minor and finally manages to hoist itself up to the proper mediant major in mm. 36–37 (although further damage to S is also apparent in subsequent measures).

It may also happen that a longer stretch of caesura-fill, branching out from the tonic authentic cadence at the end of TR, is called upon to accomplish the modulation to the new key. This is uncommon in eighteenth-century works but turns up occasionally in works in the nineteenth century, as in the first movement of Schubert's Symphony No. 8 in B Minor, "Unfinished," D. 759, first movement—i:IAC MC at m. 38 (perhaps interpretable as a i:PAC), with a modulatory caesura-fill leading to an S-space in G major that begins in m. 42.[8]

---

7. Haydn often gives us an emphatic V:PAC at a point where it is difficult to decide what its intended function might be: EEC? MC? witty or purposefully "difficult" gesture? Is its very point to place us in an ambiguous interpretive position? Or might it be the articulation of a different (third?) type of exposition altogether—perhaps one customized by Haydn for individual use or perhaps one known to him from more local traditions? (Similar situations crop up also in midcentury sonatas of less well-known composers) However one regards it, this V:PAC option (not too far past the midpoint of the exposition) is one of the important features of Haydn's conception of sonata form.

8. Related instances would include modulatory CF passages following a purposefully "wrong-key" MC, such as that in first movement of Schubert's Piano Sonata

The Medial Caesura: Common Characteristics

Within Allegro compositions (first movements and finales, most overtures) the medial caesura is often the final moment of articulation following one or more measures of preparation on a prolonged structural dominant (dramatically sustaining the earlier arrival of a I:HC or V:HC). A common sequence of events is: (1) the initial stages of TR, by and large consistently gaining in energy; (2) the attaining of the structural dominant by means of a half-cadence arrival (usually either a V:HC or a I:HC—or the minor-mode-sonata equivalents), which is frequently then locked onto as a literal or implied pedal-point (*structural-dominant lock*); (3) the prolongation of this still-active V (V_A) and the rhetorical drive to the medial caesura—a drive that sustains or even increases the energy accumulated thus far; (4) the articulation of the MC proper, the terminal gesture of the entire process. Example 3.1 shows the opening of the first movement of Mozart's Piano Sonata in D, K. 284: TR begins in m. 9; the half-cadence arrival and the dominant-lock (I:HC), holding fast onto that V_A, occur in m. 17; the second-level default MC (I:HC MC—the concluding gesture of a prolonged half-cadence process) is articulated in m. 21; S ensues in m. 22.

Not all Allegro compositions articulate all four of these events. It is possible—for a less rhetorical effect—to sound the MC at the moment of the arrival of the half cadence, thus omitting the structural-dominant lock altogether. This would be an instance of a nonelaborate, straightforward articulation of the MC. Generally considered—and if not overridden by other evidence—if TR produces a notable HC that is immediately followed by an "acceptable" new theme in the proper subordinate key, that HC may be interpreted as a medial caesura. (Put another way, the situation is staged as if the apparent S-theme has "understood" that HC to have been one.) An example may be found in the first movement of Mozart's Piano Sonata in C, K. 309, in which TR begins in m. 21 and the structural dominant is reached only with the arrival of the V:HC MC in m. 32. In addition, many slow movements—generally lyrical movements—omit the dominant-lock. The MC-effect in these gentler movements is often produced by a mere half cadence without much additional rhetorical emphasis. (The situation in the C-major Andante movement of Mozart's Piano Sonata in G, K. 283, is typical: V:HC MC in m. 8.) It may be that any prolonged dominant-lock in slow movements was intended as an unusually strong or expansive gesture.

Normally, however, in Allegro movements, in order to function as a normative medial caesura, the *forte* half-cadence arrival within TR must be additionally reinforced. The whole process often proceeds as follows:

1. The structural dominant (the half-cadence arrival, which typically precedes the MC, sometimes by several measures) is often approached through a chromatically altered predominant harmony that contains $\sharp\hat{4}$. (This scale-step is reckoned in the key within which the half cadence is to be sounded. In the case of directed motion into a V:HC, $\hat{4}$ of the new key would be $\sharp\hat{1}$ of the original tonic.) This altered predominant is frequently an applied chord (V/V, V^7/V, vii^o/V, or vii^{o7}/V in root position or inversion) or an augmented sixth chord.[9] The chromatic line $\hat{4}-\sharp\hat{4}-\hat{5}$ or $\hat{3}-\sharp\hat{4}-\hat{5}$ often appears in one

in C, D. 279. Here we have an unequivocal arrival on the "wrong dominant," iii:HC (V of E minor!) at m. 37, which V_A is immediately frozen as a dominant-lock ending with a iii:HC MC, mm. 37–41. Four bars of expanded fill, mm. 41–44, accomplish the modulation to the generically proper key, G major (V), in which S then begins, m. 45.

9. According to Allen Cadwallader and David Gagné, *Analysis of Tonal Music: A Schenkerian Approach* (New York: Oxford University Press, 1998), p. 409, n. 23 (referring to a situation in the first movement of Beethoven's Piano Sonata in G, op. 49 no. 2, discussed on pp. 311–29), when the half cadence arrival is preceded by its own applied dominant (V of V, V^6 of V, and so on), "we might refer to this goal as a *tonicized half cadence*." While accurate, this terminology might be potentially confusing. This surely means only that an unequivocal half cadence is locally supported by its own dominant. One continues to experience that dominant arrival as an active V, as V_A, not as a tonic (V_T), particularly in "lighter" cases of a mere $\hat{4}-\sharp\hat{4}-\hat{5}$ or $\hat{3}-\sharp\hat{4}-\hat{5}$ motion in one of the outer voices—where any sense of "tonicization" in the normal sense of the word is virtually nonexistent. Thus there can be no claim that anything like a "full" tonicization of that V_A has occurred: the arrival on V is still a half cadence, not a concretized

of the outer voices ($\hat{4}-\sharp\hat{4}-\hat{5}$ as the bass-line approach to V_A is especially characteristic); if an augmented sixth chord is employed, the typical bass line is $\flat\hat{6}-\hat{5}$ and the $\sharp\hat{4}-\hat{5}$ move occurs in the upper voice.[10] The texture at this moment is vigorous, highly active; the dynamics, usually a strong *forte*, will persist or even gain in intensity in the subsequent drive to the medial caesura.

As mentioned earlier, within major-mode sonatas it is not uncommon to encounter intermixtures with the minor mode, perhaps even a shift to the minor mode—usually that of the new key—in the vicinity of this half-cadence arrival or dominant-lock point, although, if this occurs at all, it may also take place earlier or later in TR, perhaps even persisting through the MC itself. An appearance of the negative minor mode participates in the generic expectation of the intensification process, either enhancing it or engaging in some other kind of dialogue with it. The mixture with minor may suggest the introduction of uncertainty, doubt, or peril into the narrative thread—or (as sometimes in Beethoven) the onset of a grim struggle in the production of the MC and S. Following such a modal shift, the ensuing *major-mode* S emerges with a sense of brightness and relief.

2. Once attained, the structural dominant is frequently prolonged, perhaps by neighbor motion, as part of the drive to the MC proper (*dominant-lock*). This may involve alternating V with a neighboring $\frac{6}{4}$, producing $\frac{5}{3}\text{-}\frac{6}{4}\text{-}\frac{5}{3}$ neighbor motion. Sometimes the neighboring $\frac{6}{4}$ is supported by $\hat{1}$ in the bass, creating an apparent V–I–V alternation. The larger point, though, is that the V_A of the half-cadence arrival is vigorously seized onto, frozen in place, kept alive by means of a specialized pedal-point effect that announces that TR is ending with a continued push toward the MC. The sense of an "HC-moment" is not released and left behind as a mere past event—as happens with most other kinds of HC phrase endings—but rather is held onto, brandished as an achievement, sustained as a continuing function with a specific role to play at this point in the form.[11]

3. The normative, unflagging drive in the space between the locking onto the structural dominant and the actual articulation of the MC is of paramount importance. Any attenuation of dynamics here should be viewed as countergeneric, or perhaps—especially by the later eighteenth century—as a less common, second-level default that calls attention to itself and challenges the prevailing norm of energy-gain. Depending on the circumstances a dynamic collapse in this space might represent the staging of a momentary crisis of confidence in one's decision to enter S-space. The S that follows the dynamically underprepared MC might

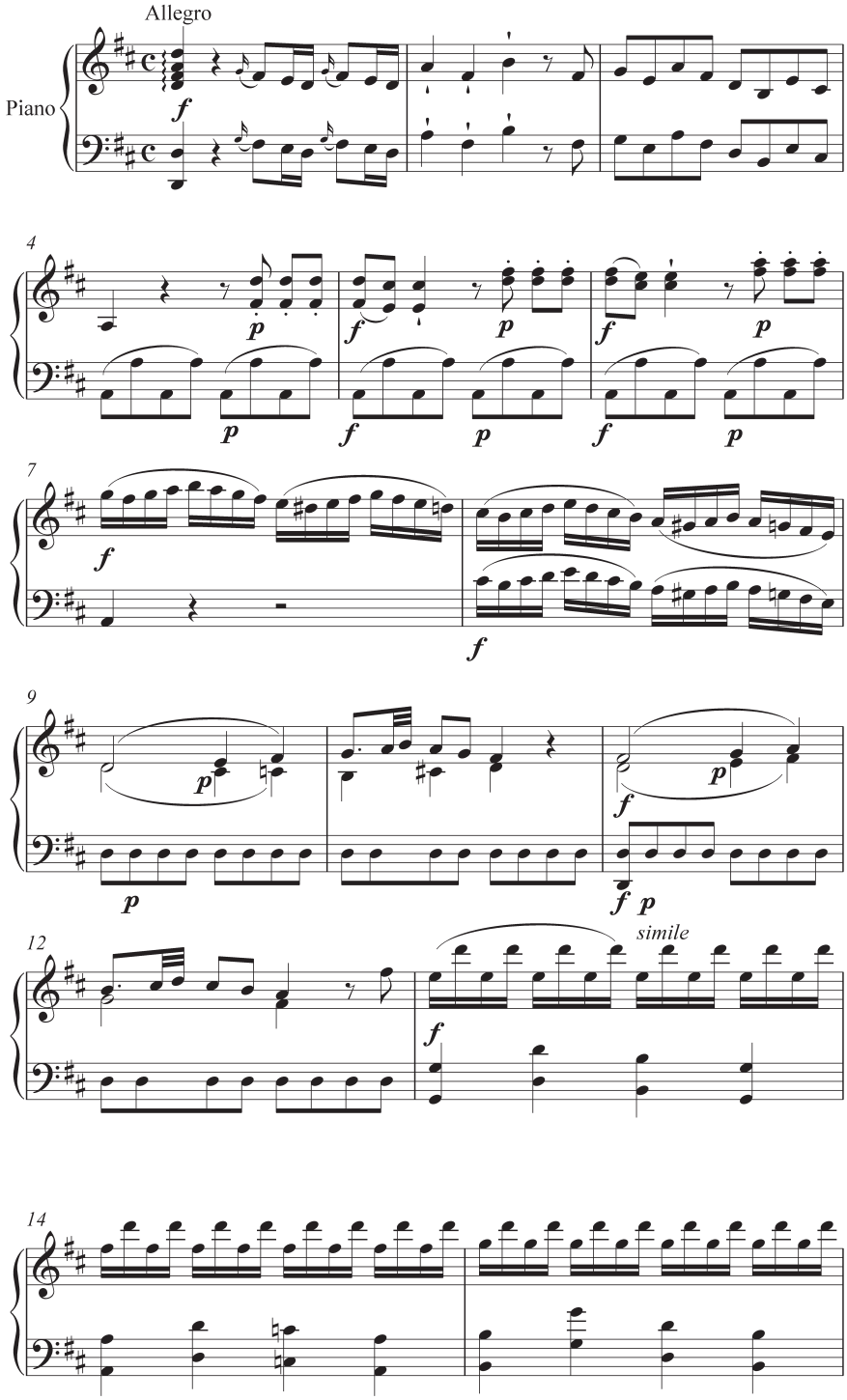
---

tonicization of that V. The difficulties of interpretation increase, however, as the sense of that "tonicization" becomes increasingly intense with differing strengths of applied dominants to the V_A. (It is possible to imagine a continuum of differing applied-chord strengths, for instance, that ultimately lead one to cases that appear to articulate a V:PAC as the MC.)

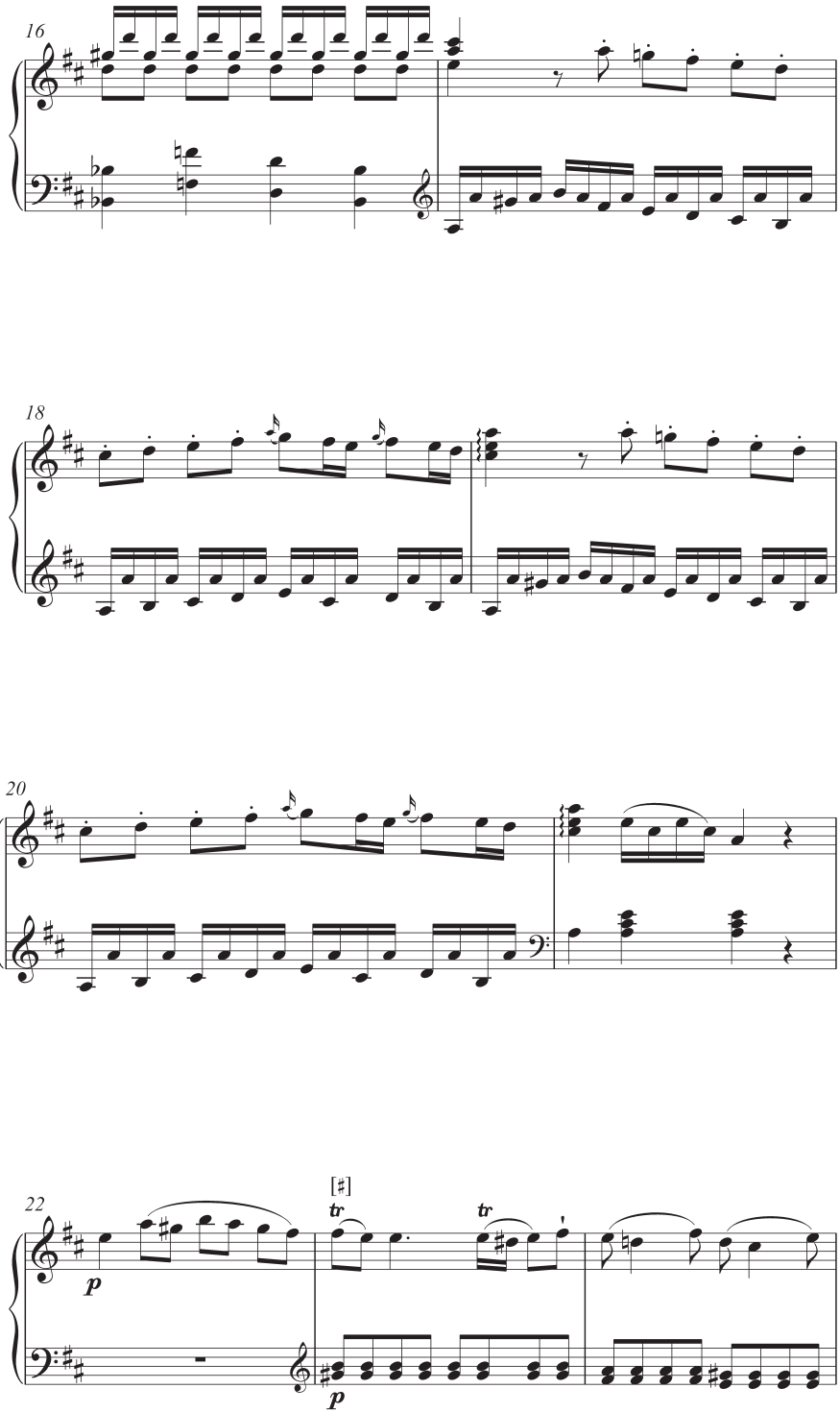
10. For a stronger sense of rhetorical emphasis it is possible to approach the structural dominant—or half-cadence arrival—more than once in fairly rapid succession. Thus once the structural dominant is sounded, it can be immediately re-sounded through energetic reiterations of the half cadence. The music can go through the cadence several times, reapproaching and rearticulating it, helping to produce the rhetorical drive toward the MC proper. The touchstone example occurs in the first movement of Beethoven's Piano Sonata in F Minor, op. 2 no. 1, where the predominant $\hat{4}-\sharp\hat{4}-\hat{5}$ bass motion preceding the half-cadence arrival is first stated in mm. 15–16, then restated twice more, mm. 17–18, 19–20. (M. 20 is the MC proper, but, unusually, it also initiates a dominant-lock and the onset of an S^0 (or $S^{1.0}$) theme, over that V. Such S-theme beginnings are discussed in ch. 7.

11. Thus our view differs from that of Caplin ("The Classical Cadence," passim, but see especially pp. 89–91, 98–100, 108–12), who, as mentioned above (n. 6), regards this dominant-lock as "postcadential" in function. We find it preferable to think of this stretch of music not so much as existing "after" the half-cadence arrival (which in a literal sense, it does) as keeping that arrival alive, refusing to let it go, animatedly spreading its sense of being still active over several more measures. Put another way, the dominant-lock may be considered a special prolongational technique that extends and "holds in place" the HC-arrival effect for a specific rhetorical purpose. See also n. 14. Caplin's English-language term for such harmonic locks is "standing on the dominant," a translation of Erwin Ratz's "stehen auf der Dominante," which had appeared in the latter's *Einführung in die musikalische Formenlehre* (3rd ed., enl., 1973). (See Caplin, "The Classical Cadence," especially p. 90, n. 101; and Caplin, *Classical Form*, pp. 16 ("postcadential standing on the dominant"), 75, 77–78, 144–45.)

EXAMPLE 3.1 Mozart, Piano Sonata in D, K. 284, i, mm. 1–24



EXAMPLE 3.1 (*continued*)



compensate for this enervation (as is suggested also at the end of no. 6).

4. At the point of the MC proper—the whole passage's terminal gesture—one often hears several *forte* hammer-blows (three is the most common number; two are also not infrequent) that ostentatiously reiterate the final dominant chord.[12] The hammer-blow effect is a common means of simultaneously bringing the energy-gain of TR to a terminal peak and beginning to discharge that tension for the subsequent drop to S (see nos. 5 and 6). The first hammer blow typically falls on a strong beat, normally on a strongly accented measure of hypermeter. The triple-hammer-blow effect may be disguised or embellished through inner subdivision of the relevant beats. (See also the similar variant in example 3.1, m. 21.) Particularly characteristic is the disposition of the hammer blows—when they are present—in such a way that the second (or second and third) is sounded an octave below the first. Within melodic phrase endings Koch referred to this formulaic octave-drop gesture (on weak beats or measures) both as a type of *Nachschlag* (a "striking afterward"—although clearly all second and third hammer blows are also a kind of *Nachschlag* regardless of octave disposition) and as a "Cäsur" that has been provided with an "Ueberhang" (an "overhanging") or "einen weiblichen Ausgang" ("a feminine ending").[13] Ascending octave leaps—presumably more energetic and expectant than descending ones—are also possible as part of this *Nachschlag* figure.

5. At the point of the MC one frequently encounters a general pause (GP), or rest, in all voices. This is one of the hallmarks of an unequivocal MC (the word "caesura" means a pause or a break in the texture), and it signals the precise arrival of the medial caesura. The silence of the caesura-gap is a watershed moment relinquishing the preceding drive and energy-gain: it articulates and represents *energy-loss,* thus initiating, usually, the subsequent drop to *piano* for S. From the vantage-point of TR the point is that a higher level of activity and energy has been now attained: the gears have shifted, and we are now prepared to enter the next stage of the exposition. In the normative mid- or late-eighteenth-century style this GP-gap typically lasts for only a beat or two (a "quick breath")—perhaps for a bar, but rarely longer. Moreover, it is certainly possible—even normative—to fill this brief gap with sound, perhaps a held note or simple scalar connective figure in one voice: we call this common procedure *caesura-fill*. This is discussed in a separate section below.

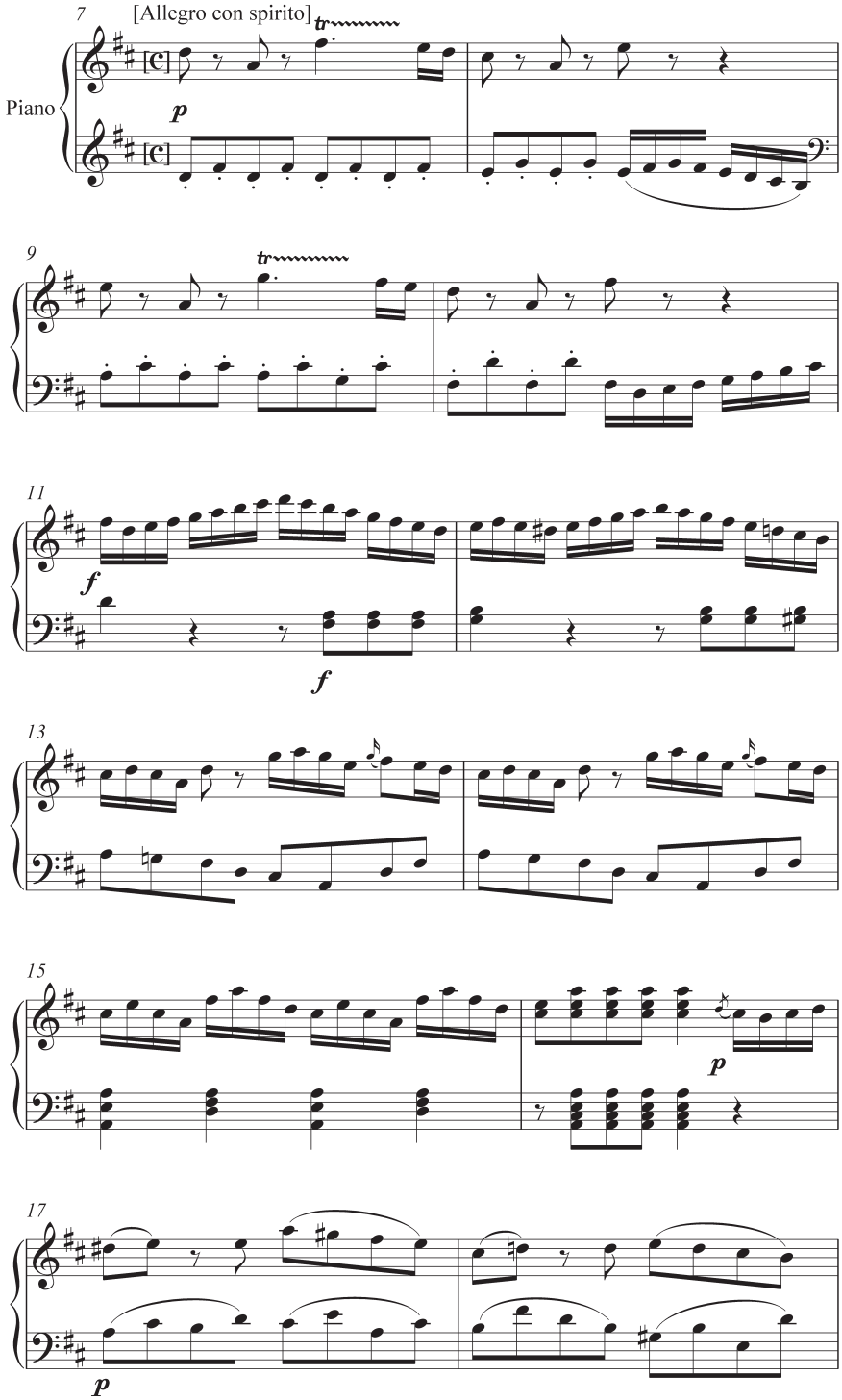
When S begins with an upbeat, that upbeat might occupy the implied GP-gap. In other words, the MC will be sounded normatively, but at the precise moment that one would expect a gap of silence, the upbeat for S ensues, dovetailed into the GP-gap. Example 3.2 provides an illustration, the TR and MC of the first movement of Mozart, Piano Sonata in D, K. 311. Following a half-cadence arrival and dominant-lock in m. 13 (introduced via a normative $\hat{4}-\sharp\hat{4}-\hat{5}$ motion in the bass), the I:HC MC occurs normatively in m. 16 (with implied hammer blows on beats 1, 2, and 3). Instead of encountering a rest (GP-gap) on the fourth beat of m. 16, we find that S begins (*piano*) with an upbeat to m. 17.

Toward the later decades of the eighteenth century (and even more so in the nineteenth) composers began to explore the effects of widening that caesura gap—opening it to a span of three, four, or more bars—and filling it with connective caesura-fill (representing energy-loss) that might serve a variety of expressive purposes. At first this widening may have

12. Within the mid- and late-eighteenth-century style the triple-hammer-blow gestures (and variants thereof) are formulaic markers of important points of structural articulation. When they are present in the exposition, they are most likely to occur as indicators of one or more of three crucial spots: the exposition's beginning; the point of the MC; the conclusion of the exposition. Needless to say, they often also appear in the same spots in the recapitulation: hence midcentury and even late-century pieces often end with references to the triple hammer blows.

13. Koch, *Versuch einer Anleitung zur Composition*, part 2, p. 394 (within subsection 95 and subsequent subsections, which deal with the central concept of the caesura). See also Koch, *Introductory Essay on Composition*, trans. Nancy Kovaleff Baker, pp. 23–24 (the translation used here).

EXAMPLE 3.2 Mozart, Piano Sonata in D, K. 311, i, mm. 7–18



been considered a deformation (a purposeful, significant distortion or overriding of the norm), but it soon became a common option (though still an expressive one) within the style.

6. Immediately following the MC proper (after the implied or actual GP-gap), one expects to find the launching of a characteristic secondary theme (S)—which may exemplify any of a number of types. (See chapter 7.) One of the most common types features a sudden change of texture after the MC-point, usually combined with a precipitous drop from an energetic *forte* to *piano* and the unfolding of a melody articulating the second expositional key. This abrupt dynamic/textural change suggests the immediate emergence of a normative rhetorical candidate for S-status (the onset of the second part of the exposition), an emergence that confirms the MC-status of the preceding HC. Particularly in large-scale compositions, this criterion is crucial: the change of texture and/or dynamics functions as a standard gesture that accepts and ratifies the preceding caesura as the MC. Refusing to initiate any of the characteristic opening types of an S-theme at this moment may signal that the preceding, proposed MC is being declined by subsequent events.

But to this general norm of the *piano* S, probably the most standard option, one should add a word of caution. Although it cannot be maintained that the beginning of an S-theme can never be articulated at a *forte* dynamic level—bustling or energetic S-ideas are especially common in midcentury orchestral works—in the context of the later eighteenth century such suddenly blurted or surging S-themes are almost invariably reactive to some earlier complication in the TR zone (especially to a complication in the MC, one type of which is suggested in no. 3 above). To be sure, the unusual, *forte* S may be found to great effect here and there in Mozart and Beethoven, but it seems to have been of special interest to the mature, ever-inventive Haydn, in whose works the S-ness of the *forte* theme, when it occurs, is usually identifiable through its monothematic incipit, recalling P. (It is encountered in several of the "London" Symphonies, with locally clever implications, as in the first movement of Symphony No. 99 in E flat, m. 48.)

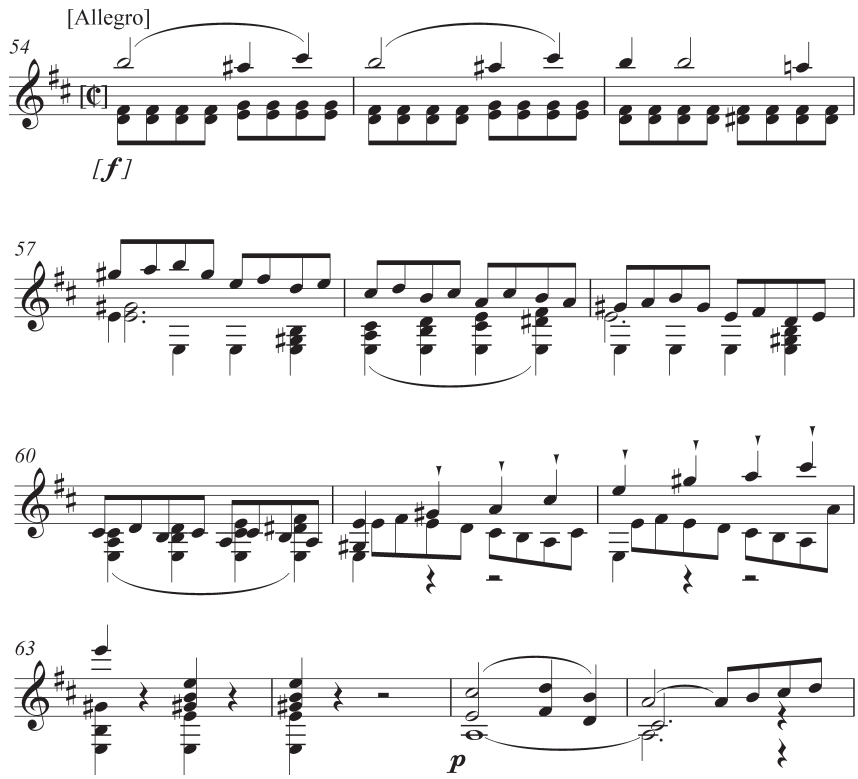
As another illustration of these principles, the V:HC medial caesura in the first movement of Haydn's Symphony No. 104 (example 3.3) is reinforced by conditions 1 (approach to the dominant through V^6_5/V, with $\hat{4}-\sharp\hat{4}-\hat{5}$ in the bass, mm. 56–57), 2 (prolongation of V_A by $\frac{5}{3}\text{-}\frac{6}{4}\text{-}\frac{5}{3}$ neighbor motion, mm. 57–62) fortified by a constant energy-gain up to the MC, 3 (unflagging energy-drive to the MC), 4 (three hammer-blows, mm. 63–64), 5 (general pause, m. 64), and 6 (change of texture and emergence of the new key, m. 65). The III:HC medial caesura in the first movement of Mozart's Symphony No. 40 in G minor, K. 550 (example 3.4) is also bolstered by these six features. Here the dominant (the half-cadence arrival) is approached through V^7/V (mm. 34–37), the neighboring $^6_4$ motion is expanded to include vii^{o7}/V over a dominant pedal, and there are only two hammer blows (m. 42, including the characteristic octave drop). Because these two medial caesuras are reinforced by the same conditions, they may be heard as roughly equivalent in strength.

Similarly, one might recall the I:HC (second-level-default) medial caesura in the first movement of Beethoven's Symphony No. 1 in C, op. 21, m. 52. Here we find the GP gap, m. 52 (two quarter rests); the triple hammer blows leading up to it, mm. 51–52; the preparatory (pre-MC) $\hat{4}-\sharp\hat{4}-\hat{5}$ motion (in the tonic) in the bass, mm. 44–45; the alternation of V with a neighboring $^6_4$, mm. 45–51 (here also anticipating the hammer blows); the drop to piano dynamics for S at m. 53. Even though we are confronting a second-level-default MC (one that marks the endpoint of a nonmodulating TR), this MC may also be considered a paradigm: it is normative in just about every way imaginable.

The Deployment Sequence of Medial Caesura Options

Another issue surrounding the identification of a medial caesura is its temporal (proportional) appropriateness—its precise placement within an exposition. This is complicated by the fact that an MC (including the possibility of a third-level default, V:PAC in major-mode works)

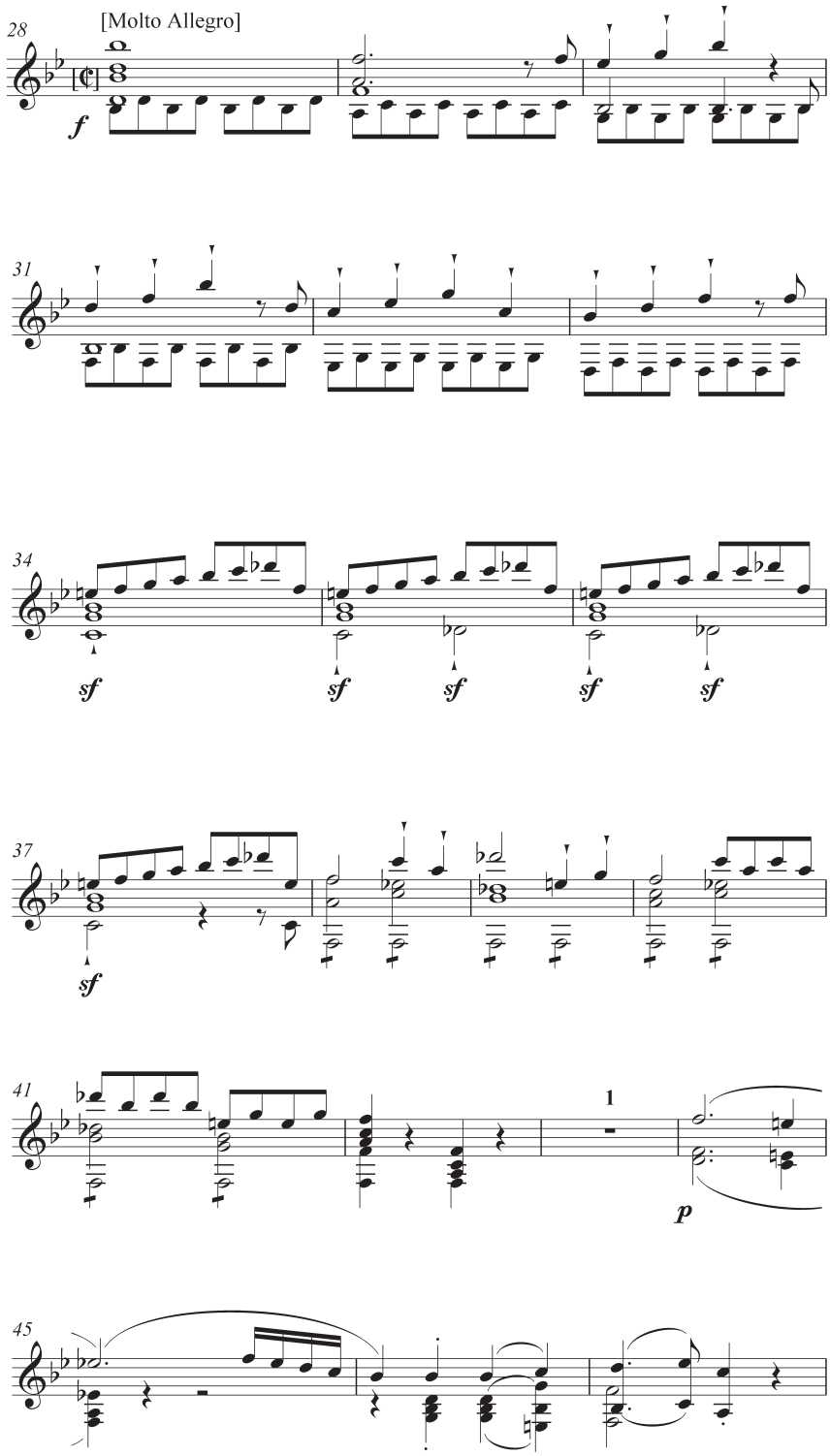
EXAMPLE 3.3 Haydn, Symphony No. 104 in D, i, mm. 54–66



could occur anywhere from about 15 to 70 percent of the way through an exposition. To be sure, this is a broad expanse of expositional space, even though most cases fall before the halfway point. Our research suggests that the deployment of the I:HC MC is flexible, occurring typically within the 15–45 percent range. Noteworthy here is the early availability of the I:HC MC. Beyond the 45 percent point—and especially in grand-scale works, such as symphonies, often earlier than this—the I:HC MC seems to have been considered either eclipsed or increasingly and rapidly left behind as a practical option. This reinforces our earlier observation that the I:HC MC was appropriate for shorter works, and indeed for more modest works it may be a more commonly selected option than the V:HC MC.

We have proposed, however, that for most analysts the conceptual reference is likely to be more ambitious in moderate- to large-scale movements, within which it is more accurate to regard the I:HC MC as the second-level default (after the V:HC MC). This results in the seemingly paradoxical situation in which the second-level default I:HC MC (which only means the one less commonly selected) is the first temporally available MC-deployment option. Some expositions even take up this norm of temporal MC availability as part of their compositional strategy. An exposition, for example, might make an early feint toward the I:HC option (by seeming to move toward or even onto the relevant structural dominant) only to renounce it or pass it by in order to produce a later V:HC or V:PAC MC. In this situation, when a I:HC is actually articulated *en route* as a seeming point-of-arrival and a dominant-lock begun on that V_A—as though it were charging normatively toward an imminent I:HC MC—such an attained lock would have to be shaken off (or "unfrozen") with a decisive plunge into further harmonic activity that now leads the music toward a modulation and a quite different

EXAMPLE 3.4 Mozart, Symphony No. 40 in G Minor, K. 550, i, mm. 28–47



MC option. A classic instance may be found in the first movement of Mozart's Quartet in B-flat, K. 172. Standard $\hat{4}-\flat\hat{4}-\hat{5}$ motion in the bass (on the last beat of m. 17) leads to what at first seems to be an "attempt" to sound a I:HC structural-dominant-lock in mm. 18–22 (albeit one decorated through oscillations with its own V^6_5 in mm. 19 and 21). A set of largely parallel, descending $^6_3$ chords, enhanced with 7–6 suspensions, shakes loose of this dominant in mm. 23–25 and moves the music instead toward an abruptly produced V:HC MC in m. 26.[14] This witty effect, found also in many of Haydn's "Paris" and "London" Symphonies, can be that of demonstrating the compositional options that the composer is choosing *not* to deploy ("No! We *won't* use this I:HC MC option! Let's select something else instead! Onward!").

The normally available range for the more common, first-level-default V:HC MC overlaps broadly with that of the I:HC MC but in general occurs slightly later. When selected, the V:HC MC option is typically placed from about 25 to 50 percent (more rarely, 60 percent) of the way through the exposition.[15] Again, this suggests that the choice of either a V:HC MC or a I:HC MC also served to predict the proportions of the remainder of the exposition and hence of the remainder of the work. I:HC MCs promise more compact works; V:HC MCs usually lead us to expect broader structures. The third-level default V:PAC MC is located in the 50–70 percent (very rarely, 75 percent) range. This is the last available deployment option, and, as mentioned earlier, it is sometimes encountered as a recovery from the staging of a failed attempt at producing a V:HC MC. Any strong caesura falling outside these boundaries is either an exceptional MC (in which case a cogent argument on its behalf would have to be offered) or, more often, no MC at all.

The precise percentage numbers admit of exceptions. More important than exact figures is the overarching principle of the normative deployment sequence of potential structural dominants and/or MCs. The initially available I:HC possibility soon overlaps with and eventually gives way to the V:HC option. If the V:HC option is not selected, the last chance to produce a two-part exposition resides with an appropriately placed V:PAC MC. Any relatively late V:PAC MC brings with it structural complications and potential ambiguities: Is it an MC or is it better regarded as the EEC?

The deployment sequence for major-mode sonatas may be represented diagrammatically (and roughly) as in figure 3.2. In summary, the important points to observe are: (1) the second-

14. In K. 172/i the emphatically sentential character of mm. 18–26—with the move away from the V_A of I only in the sentence's continuation—also contributes to this effect. Observing that mm. 18–26 are structured as a sentence, however, once again addresses the issue of Caplin's description of such post-HC dominant-locks as merely postcadential. (See the preceding remarks in nn. 6 and 11.) In K. 172/i such an assertion can lead one into contradictions, because, by Caplin's definitions, the presentation of a sentence is essentially precadential. In other words, by this logic what is claimed to start out postcadentially here—the initial dominant-lock—must also be regarded as simultaneously functioning precadentially as a sentence presentation, although in the case of K. 172/i one might observe this only in retrospect. Our mode of approaching such questions is to remain flexible, to realize that generically "predicted" behavior can and often does "change its mind" *en route* within this most mercurial of musical styles. One could even imagine other instances in which a dominant-lock that remains "locked" could be structured as a sentence—once again inviting the precadential-postcadential contradiction within Caplin's system.

15. From time to time Haydn's drive for unpredictable or non-normative originality produces extraordinary exceptions to this principle. What appears to be the V:HC MC in the first movement of Symphony No. 82 in C ("Bear"), for instance, occurs in m. 69, 68 percent of the way through the exposition. In part this occurs as a result of Haydn's earlier prolonged dalliance with the I:HC option. Jens Peter Larsen, "Sonata Form Problems," *Handel, Haydn, and the Viennese Classical Style* [orig. publ. as "Sonatenform-Probleme," in *Festschrift Friedrich Blume zum 70 Geburtstag*, ed. Anna Amalie Abert and Wilhelm Pfannkuch (Kassel: Barenreiter, 1963)], trans. Ulrich Krämer (Ann Arbor: UMI, 1988), p. 274, takes the following *piano* theme, m. 70, to be an archetypal example of the opening of the third part of a "three-part division of the exposition" (what we call a continuous exposition—the exposition-format discussed in ch. 4). In this case we disagree. Based both on the peculiar rhetorical narrative produced in TR and on the acceptably S-like rhetoric of the *piano* theme at m. 70, anticipated in the preceding drive to the MC, we believe it preferable to understand that theme as an extraordinarily late S.