

Beyond the Sonata Principle

JAMES HEPOKOSKI

One of the perennial concerns of musicology and music theory is the analysis of foundational instrumental compositions from the decades around 1800—works by Haydn, Mozart, Beethoven, and Schubert. Dealing with this music adequately presents substantial challenges. Apart from the vastness and diversity of the repertory, the bibliography of individualized approaches to it has been enormous. From the outset we are confronted with differing methodologies, entrenched interests, and conflicting analytical claims. Within English-language analytical traditions, the past three decades of the last century saw the advance of competing, sometimes complementary, modes of orthodoxy in dealing with this music. Among the most frequently embraced hermeneutic genres were: (1) the general argumentation found in Charles Rosen's two influential overviews, *The Classical Style* (1971–72, rev. 1997) and *Sonata Forms* (1980, rev. 1988); (2) the motivic quest for coherence and “unity,” typically seeking to demonstrate the generative, nonformulaic unfolding of structural shapes and contrasting ideas out of a few germinal cells presented near the opening of a piece (Arnold Schoenberg; Rudolph Réti; Hans Keller); (3) the historical-rhetorical approach, grounded in the language of the late eighteenth century and promoted in such texts as Leonard G. Ratner's *Classic Music: Expression, Form, and Style* (1980); and (4) Schenkerian analysis, anchored in such studies as Heinrich Schenker's *Free Composition* (*Der Freie Satz* [1935], English translation 1979).¹ Even today most

1. These were by no means the only hermeneutic genres in general application during the last third of the twentieth century. One thinks, for example, of the impact, especially before about 1980, of Jan LaRue's *Guidelines for Style Analysis* (New York: Norton, 1970; 2d ed., Warren, Mich: Harmonie Park Press, 1992); of William S. Newman's *Sonata in the Classic Era* (New York: Norton, 1963; rev. ed., 1972); or, more generally, of the impact of the writings and style earlier in the century of Donald Francis Tovey, whose voice and conclusions echoed through several later writers from the 1960s to the 1990s. The Schoenbergian tradition is conveyed in brief in Arnold Schoenberg, *Fundamentals of Musical Composition*, ed. Gerald Strang and Leonard Stein (London: Faber, 1967). Cf. the once-influential Rudolph Réti, *The Thematic Process in Music* (New York: Macmillan, 1951). Enriching the matter considerably at century's end is William E. Caplin, *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven* (New York: Oxford University Press, 1998).

[*Journal of the American Musicological Society* 2002, vol. 55, no. 1]

© 2002 by the American Musicological Society. All rights reserved. 0003-0139/02/5501-0003\$2.00

analyses of a “sonata-form” movement will invoke one or more of these analytical modes.

But although hermeneutic orthodoxies grow in cultural prestige and conviction—and although these four approaches are distinguished indeed—little in the analysis and interpretation of the “classical” repertory is self-evident. From time to time it proves useful to unsettle the assumptions of the various methodologies, particularly when some of those axioms become so ingrained as to pass unnoticed. One of the most deeply rooted of these within English-language scholarship is the “sonata principle.”² For the most part, this persists today as a postulate that is alluded to but concretized in varying ways and different strengths by different writers. It is rarely if ever brought forth as a sustained topic in itself. To select a recent example of its invocation, consider Ethan Haimo’s study *Haydn’s Symphonic Forms: Essays in Compositional Logic* (1995):

I take as my cue, my starting-point, a remark made by Edward Cone in his much-admired book *Musical Form and Musical Performance* . . . what he termed the “sonata principle.” . . .

The sonata principle constitutes an essential component of Haydn’s thought. It is no compositional prescriptive . . . nor does it suggest a fixed form, a framework to be filled in with details. But it is a remarkably suggestive idea, describing a basic principle of Haydn’s approach to form: highly etched statements outside of the tonic, toward the beginning of a movement, create a formal imbalance that needs to be corrected by the restatement of that material in the tonic toward the end of the movement.³

With its limitation to tonic restatement only, Haimo’s wording of the principle is a slightly altered version of Cone’s original formulation from the late 1960s. Moreover, his reconstruction of the idea is less stringent (“no compositional prescriptive”) than those of some of his predecessors, especially since it is also tempered through an interplay with four other principles (unity, [avoidance of] redundancy, variation, and [responses to deviations from] normativity), any of which, under certain conditions, might be capable of trumping it.⁴

Such a strategy is familiar in recent writing. Typically one finds a reference to the principle as something well known and long regarded as central to any sophisticated discussion of sonata form, followed by a personalized paraphras-

2. The Anglophone backdrop of both the term and the essential idea is also underscored in the useful essay by Hans-Joachim Hinrichsen, “Sonatenform, Sonatenhauptsatzform” (1996), in *Handwörterbuch der musikalischen Terminologie*, ed. Hans Heinrich Eggebrecht (Stuttgart: Steiner, ca. 1972–), 18–19.

3. Ethan Haimo, *Haydn’s Symphonic Forms: Essays in Compositional Logic* (Oxford: Clarendon Press, 1995), 3–4.

4. *Ibid.*, 3–9. It is probably significant that the sonata principle was selected to be the first of the five outlined here, perhaps the *primus inter pares*. Still: “Interestingly enough . . . the five principles are often incompatible. . . . Although the principles outlined above may be simple in theory, they are anything but simple in application” (p. 9).

ing of it flanked by one or two caveats. In the updated entry “Sonata Form” in the second edition of *The New Grove Dictionary of Music and Musicians*, James Webster reaffirmed, “The ‘sonata principle’ (Cone; the term is misleading, insofar as this is only one of several relevant principles) requires that the most important ideas and the strongest cadential passages from the second group [of the exposition] reappear in the recapitulation, transposed to the tonic.”⁵ Here we find a more limited formulation than that encountered in Haimo—only second-group themes within expositions and recapitulations are cited (surely with the implication of also including any later “cadential-closing-group” themes),⁶ not other possible themes “toward the beginning of a movement”—along with a few qualifying phrases for nonconforming cases. How does one determine what “the most important ideas” are? What may be meant is that any second- or closing-thematic module that does not reappear in the recapitulation is *ipso facto* not to be tallied among “the most important ideas,” but this would be only a circular, self-reinforcing assertion. In any event, are not all musical ideas within expositions important? Later in the *New Grove* article, Webster returned to the idea as a potentially binding principle but again added a crucial nuance in confronting works by, especially, Haydn, Mozart, and Beethoven:

In recapitulating the remaining material of the exposition, the sonata principle always applies: material first presented outside the tonic is repeated in the tonic. But in Haydn, Mozart, and Beethoven these repetitions are never mechanical. Their regular procedure was to repeat the main theme, to refer to the other first-group themes and the transition with an appropriate harmonic orientation, and to repeat the chief ideas from the second group in the same order (Mozart, Symphony in E \flat K543; Beethoven, Symphony no. 5).⁷

Notwithstanding Haimo’s and Webster’s words of caution, the sonata principle commonly functions as a hazy, background truism to be largely reaffirmed, not problematized. Its relegation to background status probably

5. James Webster, “Sonata Form,” in *The New Grove Dictionary of Music and Musicians*, 2d ed. (2001), 23:688.

6. In the quoted passage and elsewhere in Webster’s entry, the term *second group* appears often to be used broadly to mean the nontonic second half of an exposition. In the closer discussion on page 692, however, Webster distinguished between “the second group proper” and, in “large-scale expositions,” a “cadential closing group.”

7. *Ibid.*, 23:693. Cf. note 6 above for the two senses of “second group.” In 1991 Webster provided a somewhat differently nuanced, broader version of the sonata principle, when he wrote that it declared “that important material and form-defining cadences first heard outside the tonic must later be ‘grounded’ in the tonic.” Here the crucial difference is its apparent nonconfinement to expositions and recapitulations, a nonconfinement that would play a role in his accompanying argument regarding Haydn (which is taken up in the next section of this essay). See James Webster, *Haydn’s “Farewell” Symphony and the Idea of Classical Style: Through-Composition and Cyclic Integration in His Instrumental Music* (Cambridge: Cambridge University Press, 1991), 49, with footnoted reference to Cone as having first “adumbrated” the principle (n. 14). Cf. pp. 82, 127.

stems from the inadequacy of its initial formulations. For some, including its original proponents, it had once served as the *primary* guideline for sonata-form construction in the decades around 1800. And yet finding exceptions to it is not difficult. In the first movement of Mozart's Symphony No. 34 in C, K. 338, the first closing idea of the exposition, measures 64–74 (a presumably “important” crescendo theme in G major, expressing a static V, grounded in a G-tonic pedal—the beginning is shown in Example 1 as an *aide-mémoire*), never recurs elsewhere in the movement: it is cast aside after the exposition. In the first movement of Beethoven's Piano Sonata in D, Op. 10 no. 3, a “highly etched,” though modulatory, theme begins decisively on B minor (vi) in measure 23 (Ex. 2). For the present purposes we need not declare whether this is a “transition-theme” (as Donald Francis Tovey concluded); the onset of a “first modulation” to V through a supposedly conventional vi, producing the momentary effect of a “three-key exposition” (as Rosen proposed, adding, “there are no surprises [here]; we merely stop on the road to V”); the onset of a “modulating subordinate theme” (as suggested by William E. Caplin); or some other functional unit.⁸ It suffices to note it as at least fulfilling Haimo's criterion—but probably not Webster's—as a “highly etched [statement] outside of the tonic, toward the beginning of a movement,” and to observe that this theme is never sounded in the D-major tonic at any point in this movement: it is absent from the development; it recurs in the corresponding place in the recapitulation beginning on E minor (ii; m. 204);⁹ and it is not revisited in the coda. In Beethoven's F-minor *Egmont* Overture, Op. 84, the secondary and closing themes are first sounded in A♭ major (III; Ex. 3a–b) but return only in D♭ major (VI) in the recapitulation, never to be heard again. Similar examples could be multiplied at some length. Each case invites individual attention, of course, but the larger point is that they suggest a provocative group of instances that do not conform to the presumed sonata principle, at least in the versions provided above.

8. My own view, using terminology to be discussed below (see also n. 10), is that the theme is the first module of a trimodular block (TMB)—a fairly common, though quite sophisticated, late eighteenth-century strategy of elaboration involving the production of apparent double medial caesuras within the exposition. In this case, TM¹ (m. 23) follows immediately after a (first) medial-caesura deformation and seamlessly incorporates a TM² function several bars further on, particularly in the approach to the authentic cadence at measure 53. (TM³—probably to be regarded as the more convincingly “real” second theme, S—emerges with the upbeat to measure 54.) See the discussion in James Hepokoski and Warren Darcy, “The Medial Caesura and Its Role in the Eighteenth-Century Sonata Exposition,” *Music Theory Spectrum* 19 (1997): 115–54 (esp. 150 on Op. 10 no. 3). For Tovey's reading, see his *Companion to Beethoven's Pianoforte Sonatas* (London: Royal Schools of Music, 1931), 56; for Rosen, see his *Sonata Forms*, rev. ed. (New York: Norton, 1988), 246–48; for Caplin, see his *Classical Form*, 119–20 (“modulating subordinate theme”), 273 n. 74 (Op. 10 no. 3/i).

9. Measures 23–30 move from B minor to F♯ minor (vi to iii), the parallel measures 205–12 from E minor to B minor (ii to vi). The two passages are obviously related by the principle of fifth-transposition, a relationship commonly found between parallel passages in the exposition and recapitulation. The issue of fifth-transposition and its relation to the sonata principle will be addressed further below.

Example 1 Mozart, Symphony No. 34 in C Major, K. 338, first movement closing theme, mm. 64–71

64 [vln.1] *pp*

68 *crescendo*

Example 2 Beethoven, Piano Sonata in D Major, Op. 10 no. 3, first movement, mm. 23–30

23 *p*

27 *p*

Example 3 Beethoven, *Egmont* Overture, Op. 84: (a) mm. 82–85; (b) mm. 105–11

(a)

82 [strings] *ff* [cl.] *p dolce* [fl.]

(b)

[vln.1] 105 *f* *sf* *sf*

What follows is by no means a wholesale rejection of the familiar sonata principle *per se*. (That some version of the idea is relevant to the vast majority of sonata-form compositions of the period is obvious.) Rather, it is an attempt to call further attention to its limitations, along with the ways in which it has been imprecisely laid out or misapplied in influential writing. This entails the investigation of two interrelated matters. On the one hand, it is helpful to ponder the original texts that foreshadowed or first elaborated it as a guideline. What were the claims either explicitly made or strongly implied in those initial formulations? To what extent were these claims accurate or reliable? While this aspect of the present essay, concentrated especially in its first half, requires a close reading and testing of these texts, its purpose is not that of mere critique but that of clarification. Through this process we may hope to arrive at a clearer understanding of what may and what may not properly be expected along these “sonata-principle” lines: to weed out a few immoderate proposals that have been made on its behalf; to suggest the range of applicability of what remains.

On the other hand, the second half of the essay tilts toward a somewhat different concern relating to the reception history and subsequent applications of the concept. There I seek to illustrate that analytical appeals to the sonata principle (or to workable adaptations thereof) can be used to provide inappropriately quick solutions to unusual sonata-form pieces, especially to those that have non-normative thematic or tonal layouts in their recapitulations. Building on features from the earlier portion of the essay, I pursue this through the twinned strategies of a close reading of an existing analysis or two—opening them to questions of broader implication—coupled with the offering of an alternative, hopefully richer, mode of inquiry to deal with the same pieces. Ultimately, I suggest that the leading enticement of the sonata principle as a generalized and protean concept may be its ability to paper over difficult musical problems with a catchphrase and that its invocation sometimes does more to short-circuit analytical thought than it does to clarify it. Along the way, in these and a few related analyses, I hope to suggest other factors that might also serve to introduce a new and different mode of analytical questioning “beyond the sonata principle.”¹⁰

10. This different style of analysis (“Sonata Theory”) is elaborated in James Hepokoski and Warren Darcy, *Elements of Sonata Theory: Norms, Types, and Deformations in the Late-Eighteenth-Century Sonata* (New York: Oxford University Press, forthcoming), aspects of which have been introduced in Hepokoski and Darcy, “The Medial Caesura,” 115–54. Sonata Theory includes some still-unfamiliar concepts and terminology (labels, abbreviations, and acronyms) that will inevitably surface in what follows, although in each case I seek to provide brief, workable definitions sufficient to the present context. (The goal here cannot be to derive or justify the entire system.) Two additional points are worth mentioning. First, while there are points of contact between the present article and the forthcoming book, this essay, taking up a central issue and several examples in more detail, is not an extract from the latter. Second, this essay was conceived as one of a complementary pair of articles. Its sibling is “Back and Forth from *Egmont*: Beethoven, Mozart, and the Nonresolving Recapitulation,” *19th-Century Music* 25 (2002): 127–54.

Tovey, Ratner, Cone, and Others: Basic Texts

In sophisticated English-language writing on music, the twentieth century was marked by a long and distinguished tradition of resisting thematic labels and textbook encapsulations of sonata-form procedures. Characteristically, the dismissal was made on two interrelated grounds, one observational, the second aesthetic and/or ideological. According to the first, sonata procedures circa 1800 were more flexible and variable than any existing textbook descriptions had suggested—especially those quick digests foisted upon undergraduates and popularized in middlebrow commentary on the standard repertory. According to the second, master composers concerned with the demands of art—as opposed to formula—knew better than to fall prey to stereotyped formal patterns (in which, therefore, we as analysts ought not to shackle them). The tradition was in large part secured by Tovey, whose writings and analytical tone had an often determinative impact on many subsequent English-language discussions. What present-day analyst or music historian has not been steeped in statements similar to Tovey’s from 1927, 1929, and 1934?

The term “second subject” is, for reasons which will soon appear, the most misleading in the whole range of our British musical provincialisms: it is unknown in Germany, where the term used in its place is *Seitensatz*, a term conveying no false ideas. . . . [As for sonatas grounded in “two contrasted themes”] these things are often conveniences, and it saves trouble to make a convenience into a convention. But we need not put up with a convention that is no longer convenient.

It was not Beethoven’s forms but his dramatic power that gave him the reputation of a musical revolutionary. . . .

Musical theory has had to struggle with material hardly ever more than a generation older than the theorist; and the generic inferiority of the theorist to the creative artist shows itself in the choice of authorities for ‘classical’ procedure. If these authorities were avowed, the mischief would not be serious: students would know that ‘normal’ form is ‘normally’ exemplified only by Spohr and Hummel. . . .

But the names which orthodoxy associates with this ‘normal’ form are those of Haydn, Mozart, and Beethoven, three composers who differ from each other in their treatment of form as profoundly as they differ in other aspects of style and matter. . . . The first difficulty is to find two movements by Mozart that are sufficiently alike to produce any such uniformity of procedure as can have served Spohr’s purpose.¹¹

There are no rules whatever for the number or distribution of themes in sonata form. When critics tell us that Mendelssohn is weak “in second subjects, where the human element is required,” they disqualify themselves by a terminology as

11. Tovey, “Some Aspects of Beethoven’s Art Forms,” in his *Main Stream of Music and Other Essays* (Oxford: Oxford University Press, 1949), 271–97, at 272–73. The essay was originally published in *Music and Letters* 8 (1927): 131–55.

useless as that of the friend who did not see where the painter was going to put his brown tree.¹²

My first axiom is that the main difference between Science and art is that there is no such thing as Art with a capital A. . . .

There is no such edifice [to which the artist has a “duty” to “contribute”]. There are individual works of art, and it is the business of each individual work to be a whole. . . .

How can we discover and estimate the fundamental hypotheses of a work of art? They are not, as Cardinal Newman suggested in a famous passage, like the rules of a game, to be learnt beforehand, but they are self-explanatory results of the contents of the work. . . .

Music has no temptation to be anything but an art pure and simple; and many works of its great masters are amazingly perfect in conception and usually perfectly preserved.¹³

Tovey’s dismissal of any view of musical practice construed along norms of convention and the practice of labeling—his imperial disdain for mere analytical system—proved alluring, as did his magisterial prose style, proclaiming the attainment of a position of privileged connoisseurship, of having seen through and outgrown schoolboy orthodoxies.¹⁴ Other writers were also calling the

12. Donald Francis Tovey, “Sonata Forms,” entry for the fourteenth edition of the *Encyclopaedia Britannica* (1929), reprinted in *Musical Articles from the Encyclopaedia Britannica* (London: Oxford University Press, 1944), 210–12; reissued in 1956 under the title *The Forms of Music*. (For the fourteenth edition, Tovey thoroughly overhauled his earlier, 1911 essay, “Sonata Forms,” from the eleventh edition—which, for instance, contains no references to brown trees. While not incompatible, the two articles differ markedly in language, tone, and presentation.) The phrase “there are no rules whatever” was also reused in Tovey, “Some Aspects of Beethoven’s Art Forms,” 274.

For other statements of Tovey’s opposition to the terms *first subject* and *second subject*, see (among many other sources of this *idée fixe* in Tovey) “Sonata Forms” (1929 version and *The Forms of Music*), 209 (“the provincial fact that English musicians have fastened on the terms ‘first subject’ and ‘second subject’ instead of translating the excellent German terms *Hauptsatz* [principal member] and *Seitensatz* [subordinate member]”; *A Companion to Beethoven’s Pianoforte Sonatas*, 1 (“The listener has no business even to know that there is such a thing as a ‘Second Subject’ until he hears it”); and “Some Aspects of Beethoven’s Art Forms,” 281 (“At bar 22 the section misnamed ‘second subject’ begins”). The Germanic terminology (along, perhaps, with his translation) was not as clear-cut as Tovey implied. To be sure, A. B. Marx had employed the terms that Tovey noticed—though not rigorously (moreover, Marx’s own technical sense of “Satz” needs to be considered here)—but other terms for these thematic events, of course, had included *Thema* and *das zweite Gedanke* or *das zweite Thema* (Birnbach, 1827); and *principal subject* and *middle subject* (Czerny, 1837–49). On Tovey, see also Joseph Kerman, “Tovey’s Beethoven” (1975–76), in his *Write All These Down: Essays on Music* (Berkeley and Los Angeles: University of California Press, 1994), 155–72.

13. Tovey, “Musical Form and Matter” (originally The Philip Maurice Deneke Lecture at Oxford, 4 June 1934), reprinted in his *Main Stream of Music*, 160–82, at 160–62.

14. Paradoxically—and notwithstanding the admonitions of the book’s introduction—the reverse impression may be obtained from the analyses in Tovey’s schematic *Companion to Beethoven’s Pianoforte Sonatas*.

thematic-architectonic basis of the form into question. Thus Schenker, for instance, from 1935:

Here [in the presentation of sonata form] . . . it is necessary to discard the concepts and terminology of conventional theory. . . . It does not matter that so many designations are offered for the prolongation of the primary tone of the fundamental line (“first theme,” “main theme,” “first theme-group,” and such); what matters is that none of these designations answers the essential question, not one explains why the first prolongation takes just this particular course and no other.¹⁵

Schenker himself, along with several of his students, may have had only a modest impact on academic considerations of musical form before mid-century, but the general trend of all such arguments was clear. One result was that the scholarly suspicion of prior, reductive textbook-schemes for sonata form increased in English-language music institutions—and rightly so—with the rise of a more sophisticated modernist musicology and music theory. With this suspicion a central analytical problem began to surface. Since the treatment of sonata form in the hands of the classical masters was variable—and since writers now wished to avoid descriptions that were overly labeled or potentially constraining—was it possible to establish the principles underpinning sonata form in a more flexible way?

Different scholars came up with differing answers, some of them grounded, like Leonard Ratner’s, in a shift away from a thematic conception of sonata form toward a harmonic one—a view that by the 1960s was regarded as the more intellectually advanced position. As Ratner had famously written in 1949:

The familiar textbook outline of sonata-form gives the approximate order of material for many movements. Still, the thematic conception of classic sonata-form is by no means universally valid. . . . The bi-thematic nature of classic sonata-form has been considered axiomatic. . . . Yet this assumption of a fundamental polarity of two themes does not hold in the case of numerous movements that contain more than two themes; it also breaks down when applied to the monothematic expositions that are frequent in Haydn. . . .

. . . it is apparent that the principal items in the thematic conception of sonata-form—bi-thematicism, thematic contrast, and well-defined areas of statement and development—are not valid as basic formal criteria. Regardless of the aesthetic value of thematic material, it cannot be said that the generating factor of classic sonata-form is a fixed relationship of the themes.¹⁶

15. Heinrich Schenker, *Der Freie Satz* (1935), trans. Ernst Oster as *Free Composition* (New York: Longman, 1979), 1:133.

16. Leonard G. Ratner, “Harmonic Aspects of Classic Form,” this *Journal* 2 (1949): 159–68, at 159, 160). Cf. the useful discussion of this issue in Mark Evan Bonds, “The Paradox of Musical Form,” which anticipated some features of my discussion here (chap. 1 of his *Wordless Rhetoric: Musical Form and the Metaphor of the Oration* [Cambridge: Harvard University Press, 1991], 13–52). Cf. also Tovey in notes 11–13 above and 17–19 below.

Ratner's conclusion—"from the evidence given above, it seems clear that in the classic era, form was conceived primarily in terms of key relationships"¹⁷—resonated well with Tovey's suspicion of thematic labels (Ratner's description reads similarly to Tovey's *précis* of sonata form in the fourteenth edition [1929] of the *Encyclopaedia Britannica*),¹⁸ and it was bolstered with citations from eighteenth-century theorists. One of its far-from-negligible side benefits was its power to sweep aside the need to assess the "meaning" of potentially puzzling thematic arrangements, variants, expansions, reorderings, or omissions in the second half of sonata movements. When encountered, such anomalies could be neutralized in a stroke by enlisting them as evidence for the secondary or casual status of thematic arrangement (buttressed by the refrain that sonata form was no "fixed form") and, consequently, as evidence both for the superiority of the harmonic conception of the form and for the unpredictability shown by the masters of the style. Still, the significance of thematic manipulation, however freely encountered, could not be disregarded altogether. What was the role, if any, of themes within this harmonic plan?

The attack on the notions of "textbook form" and confining, fixed grids continued in the 1950s. It was during this decade that the term *sonata principle* seems first to have surfaced as part of the search for a postulate of order underlying the music of the master composers. Its earliest usage had nothing to do with tonality or tonal arrangements. The term (with hyphen, *sonata-principle*) appears to have been coined in an essay from 1952 by the British writer Philip T. Barford. In Barford's view, the sonata-principle was no technical device, but a manifestation of creativity and imagination, a dialectical interplay between the "objective" or "abstract rules of sonata form" and the "subjective feeling" of the stunning particulars of the individual masterly composition. Moreover, perceiving the principle was linked with devotional asseverations of aesthetic value—projecting an aura of sophistication around the term that would persist for decades. It was nothing less than a "transcendent synthesis" of universal and particular, finally arrived at historically with Beethoven. It put an individual, coherent stamp of "unity," "organic wholeness," "oneness of purpose," "romantic" imagination, and a "logical integration" of opposing contrasts onto the whole, avoiding the easy complacencies of lesser composers, who might be content with a mere "fanciful association" of contrasts poured into a preexisting "pastry-mould."¹⁹

17. Ratner, "Harmonic Aspects of Classic Form," 163. Cf. Tovey on Beethoven's Op. 22/i ("Some Aspects of Beethoven's Art Forms," 282, 283): "Our analysis has depended on two things: first the assertion of key and key relation, which is, so to speak, the topography of music, and secondly, the lengths of the phrases"; "The function of the exposition has been to assert two keys, the tonic and (in this case, as usual) the dominant."

18. Tovey, "Sonata Forms" (*The Forms of Music*), 214. See note 12 above. A parallel summary may be found in Tovey, "Haydn's Chamber Music" (1929), in his *Main Stream of Music and Other Essays*, 54–55 ("This description has avoided all assertions as to how many themes there are, and how they are distributed").

19. Philip T. Barford, "The Sonata-Principle: A Study of Musical Thought in the Eighteenth Century," *The Music Review* 13 (1952): 255–63. The motivating factors of this imaginative unity

In a book from 1957 commenting on a number of composers, *The Sonata Principle (from c. 1750)*, Wilfrid Mellers essentially endorsed this view—although the principle in question was anything but clearly defined. Once again, he invoked the flexibility of the masters and (as Tovey had done) cast aspersions on the concept of sonata form itself:

The term is misleading, because it suggests that “form” can exist independently of the musical “content.” Sonata, like fugue, is not so much a form as a principle, an approach to composition. . . . One might even say that there is no such thing as sonata form; there are only sonatas. . . . For the masters style never becomes dogma; form remains principle which perpetually renews itself under the pressure of experience.²⁰

The turning point in the history of the term *sonata principle* came with Edward T. Cone’s redefinition of it in a set of lectures delivered in January 1967 at the Oberlin College Conservatory and published in *Musical Form and Musical Performance* in 1968, a book that much impressed the generation of scholars coming of age around that time. In the third lecture, “The Picture Gallery: Form and Style,” Cone provided a few observations concerning “the formal design that is ideal for the Classical style, as the fugal or ritornello design was for the Baroque. . . . [This is,] of course, the sonata-allegro.” How was the problem of formal diversity within the sonata to be handled? As was typical of the time, thematic concerns were brought up but subordinated to tonal matters.

It is easy to see that the Classical style, with its possibilities of rhythmic variety, is admirably adapted to polythematic forms, and to the kind of treatment that typifies the development section. But more important than the form as pattern is the unifying principle behind it, which, I believe, is not to be found in its bithematicism, or its developmental aspect, or its binary or ternary (take your choice!) structure. Let us recall for a moment that the principle underlying both the fugue and the concerto was the recurrence of the theme at every important point of harmonic arrival. The corresponding principle for the Classical

within the particular were “rhythmic flow” or “smooth rhythmic impetus” in J. C. Bach and Mozart; and “development” of opposing contrasts in Beethoven. Barford’s “pastry-mould” was probably an adaptation of Tovey’s famous objection to the unimaginative conception of forms as “pre-established jelly-moulds,” as, for example, in “Brahms’s Chamber Music,” in his *Main Stream of Music and Other Essays*, 224. Cf. Tovey’s inveighing at greater length against such “moulds” in “Some Aspects of Beethoven’s Art Forms,” 288–89, 296.

20. Wilfrid Mellers, *The Sonata Principle (from c. 1750)* (Fair Lawn, N.J.: Essential Books, 1957), 3. Some of the passage above is also quoted in Hinrichsen, “Sonatenform, Sonatenhauptsatzform,” 19, which provides a brief discussion parallel to that provided in this section of the present essay. In Tovey’s (n. 11 above) and Mellers’s half-serious (?) denials of the reality of sonata form, we may recognize a latter-day instance of what medieval philosophers called nominalism: the disavowal of the existence of abstract concepts in favor of particulars. Stressing the uniqueness of the individual art object (always perceived as escaping the confines of a reductive classification regarded as fictive), this line of pronouncement would prove useful to the anti-textbook faction within the ongoing debate on sonata form.

style—let us call it the sonata principle for want of a better term—is somewhat more complex. It requires that important statements made in a key other than the tonic must either be re-stated in the tonic, or brought into a closer relation with the tonic, before the movement ends. Expressed thus, the principle covers many aspects of formal treatment. It applies, most obviously, to the role of the “second subject” in exposition and recapitulation. But it also explains why Beethoven takes such pains in the coda of the first movement of the *Eroica* to re-introduce the theme of the development, and in such a way as to modulate directly to the tonic. It suggests why Mozart sometimes introduces into the coda of a rondo the cadential tag from a central episode (as in the Andante of the Piano Concerto in E-flat, K. 482); and why Beethoven, in a similar position, sometimes makes a special bow to the key of such an episode, even though he may feel it unnecessary to mention its theme. (See, for example, the last movement of the Sonata Pathétique, Op. 13. Here, instead of referring to the theme of the central episode, the composer associates the opening theme with the key of the episode, which is at the same time brought into close relation with the tonic.)²¹

Much was claimed in these few sentences, and we shall have to spend some time with them. To begin with, we should acknowledge the casual, even impressionistic, tenor of these remarks. Cone was more concerned with sketching out a broadly generalized point than with formally proposing a group of analyses (or even a set of sentences), each of whose implications was to be held up to critical scrutiny. The public lecture in which they were first delivered—within a conservatory context—sought to provide only a humanistic overview (“a short walk through a museum”) of contrasting characteristics of “style” in the “Late Baroque,” “Classical,” “Romantic,” and “Modern” periods. Much of the lecture was devoted to matters of rhythm and its implications for twentieth-century performance. In these surroundings, the comments on “the sonata principle” seem something of a postscript to a more developed discussion of rhythm in the Classical period—the importance of meter and the measure (rather than the mere beat), the rhythmic variety of individual themes and the balanced interplay between “abstract meter” and the “concrete rhythm” within the individual exemplar, considerations of rhythmic flexibility and “real” versus “notated” measures, a brief observation concerning what we would later call hypermeter, and the like.²²

Since Cone’s perfectly reasonable objective in 1967–68 was to furnish an efficient succession of suggestive generalizations about different matters of historical “style” and to illustrate each one by pointing in the direction of an example, it might seem uncharitable over three decades later to submit his

21. Edward T. Cone, *Musical Form and Musical Performance* (New York: Norton, 1968), 76–77. The passage continues with a brief reference to the principle’s application to operatic numbers (“the Trio from Act I of *The Marriage of Figaro*, for example”), which will not be discussed here.

22. *Ibid.*, 59 (“short walk” and the periods), 72 (“real” and “notated” measures; meter and measure rather than beat), 73 (“the measures become so short that they really function as beats”), 74 (“abstract meter” and “concrete rhythm”).

observations to a rigor of investigation not fully congruent with the spirit of the original occasion. Still, in varying degrees of strength his remarks lodged deeply—and almost at once—into the network of assumptions characteristic of subsequent English-language sonata-form analysis: they have taken on a much diversified reception history of their own. This issue of the continued attractiveness of Cone’s remarks becomes all the more engaging when one realizes that several elements in them were overstated, inadvertently misleading, unclear, or simply incorrect. But one may arrive at such a conclusion only through a close reading of the text, and it is that to which we now turn.

We may first note that Cone limited the principle to apply to workings of material within individual movements only—that is, to complete, presumably closed structures within the Classical style, which were to be understood as operating under implied (twentieth-century high-modernist?) imperatives of economy and balance. (“For in a movement so constructed nothing is lost; everything that occurs will have its influence on the outcome and will have to be reckoned with before the piece is over.”)²³ It was the duty of each movement to follow its own, self-enclosed sonata principle, whatever larger organization might have been imparted to the piece as a multimovement whole. Second, at least in his initial presentation of the idea, Cone used the word “requires” with regard to the strength of the principle’s enforcement. This implies that he was conceiving it as a strong mandate. (We shall deal further below with his subsequent softening of the verb—apparently for rondos and sonata rondos—to “suggests.”) It was first offered as nothing less than the “unifying [and] underlying . . . principle for the Classical style.” Cone, in short, was most interested in demonstrating the unsuspected breadth of its reach.

He thus implied that the principle was applicable beyond sonata forms *per se*. In fact, it “covers many aspects of formal treatment.” This phrase might suggest (through analogy with the diverse genres of fugue and concerto in the Baroque period) that the sonata principle is operative in movements that are not sonatas at all. But although his formulation permits this inference, Cone could not have intended this in any broad sense. We may presume, for instance, that he did not consider the principle to cover a nontonic trio passage within a tonic-key minuet. Surely he was not arguing that the B \flat -major trio within the D-major minuet of Haydn’s Symphony No. 104 is under an imperative to be restated and resolved later in the movement. For the most part, it is likely that we are dealing with a potential applicability only to sonata-oriented (or binary-based) compositions—and to certain types of rondos as well. (This topic will be taken up in due course.)²⁴

23. *Ibid.*, 77.

24. Cf. Haimo, *Haydn’s Symphonic Forms*, 4: “Depending on the structure of the individual movement, the sonata principle can operate in works that might otherwise be described as binary (and rounded binary), aria, concerto, fugue, and rondo. However, as we shall see in more detail in the following chapters, the sonata principle does not normally function in form types (like strophic variations) that result from the concatenation of tonally complete formal units.”

Nevertheless, within sonata-form movements proper the principle was not to be confined to nontonic secondary and closing themes and their normative tonic recurrence in recapitulations. This was an arresting proposition, since the pivotal idea underlying it had been extrapolated from conventional recapitulatory treatments of expository material. (This must be why the concept was named the *sonata* principle in the first place.) This more extended proposal set Cone apart from Tovey, who had made no such claim. The latter's account of sonata form had been straightforward, modestly flexible, and almost self-evident, at least when applied to standard cases. ("Eventually a return is made to the tonic, and so to the recapitulation. This recapitulates the exposition, but it gives the second group in the [t]onic, and so completes the design.")²⁵ Cone, however, passed lightly over more obvious issues of standard recapitulatory resolutions of the "second subject" in order to produce his first illustration, which concerns a celebrated theme from the *Eroica* that is present in neither the expository layout of themes nor their recapitulatory recasting. That theme surfaces only in the movement's development and coda, both of which structural spaces, therefore, he construed as normatively operating under the sonata principle as stated. But such a claim was overextended. This is because Cone, surely unintentionally, cited the uncommon occurrence as the norm ("But it also explains why Beethoven . . ."), which had the effect of encouraging his readers to regard a strikingly original compositional decision—one to which we shall return below—as merely business as usual. Suggesting that new and "important" nontonic themes within the developmental space are to be swept under the control of the principle, such a wording draws our attention away from much evidence to the contrary. We are invited to suppose that new, nontonic, post-expositional themes in the decades around 1800 were customarily brought back and resolved at some point "before the movement ends." This is not so.

The most characteristic instances of such themes would be episodes beginning or inserted into developments. Many of these—most, in fact—do not follow the claimed principle at all. To begin with an obvious example: In the finale of Beethoven's Piano Sonata in F Minor, Op. 2 no. 1, a stormy, *prestissimo* exposition moves from tonic to minor dominant, only to lead to a development that starts with (and is substantially occupied by) a closed, lyrical episode in A \flat major, a new theme, *sempre piano e dolce*, that lasts for a full fifty-one bars, measures 59–109 (Ex. 4). The theme eventually gives way to more typically developmental material, but the "important" new theme from the development is never heard again—quite unlike the situation in the *Eroica*. One would presume that no binding principle was being violated here: we should not regard Beethoven's deployment of this A \flat episode as carelessness or a compositional mistake.²⁶ (Tovey's general description of the function of

25. Tovey, "Sonata Forms," 214. (In this source the word "tonic" is misprinted as "lonic.")

26. I interpret the presence of the A \flat rounded-binary episode—a kind of inset lyric song—primarily as an act of compensation for the bypassing of the major-median III within the exposition in favor of the more fatalistic move to minor v. In short, from one perspective this is

Example 4 Beethoven, Piano Sonata in F Minor, Op. 2 no. 1, finale, mm. 59–68

59 *sempre piano e dolce*

64

one typical kind of developmental episode seems apt here: “It may be a relief from unusually concentrated figure-work in the exposition.”²⁷ Within this context, Tovey does not seem to have considered the notion that any larger tonal claims—most notably, a need for a resolving restatement elsewhere in the movement—need be made about it.) Another, similar case of an isolated, “important” melody in the development—and here the “new” idea, like that in the *Eroica*, occurs after the development proper has begun—may be found in the first movement of Mozart’s Piano Quartet in G Minor, K. 478, measures 104–33, beginning in C minor (see Example 5, the material of which, subsequently sequenced in several nontonic keys, touches briefly, *en passant*, on a transient G minor [m. 120, surely no “resolution” of the idea], passes through that key, and never returns thereafter). The episodic developmental

something of an allusive back-reference to a secondary-theme-that-might-have-been had the exposition’s storms cleared for a major-mode space of respite. Such a proposal should not be misunderstood. I am not maintaining that the A \flat theme “is” the hoped-for major-mode secondary theme, now displaced into post-expositional space: its expansiveness and highly formalized, rounded-binary structure would seem to rule that out. The suggestion is only that, deformationally and in the wrong structural place, it looks back at an expositional space of loss, at a missed opportunity. Additionally, of course, the rounded-block character of the A \flat episode might suggest a temporary intermixture with certain common features of the sonata rondo, another typical choice for finale-construction. Still, as its expositional and development-recapitulatory repeat signs (among other factors) make perfectly clear, the movement is governed principally by the norms of “sonata form”; it is not a sonata rondo. As a consequence, the apparent “rondo-episode” character of this passage should not exempt it from Cone’s claims. On the contrary, it only makes it a more obviously “important” theme within the developmental space. (Moreover, as will be discussed below, Cone is willing to include nontonic interior rondo-episodes as passages that might be affected by the sonata principle.)

27. Tovey, “Sonata Forms,” 215. In this passage Tovey was explicating typical developmental episodes in Mozart, not in Beethoven.

Example 5 Mozart, Piano Quartet in G Minor, K. 478, first movement, mm. 104–11

space of the first movement of Mozart's Symphony No. 34 in C, K. 338, is also instructive along these lines.

Rather than multiplying instances of this situation—not a difficult enterprise—it may be useful to confront a potentially controversial pre-*Eroica* example and the complex claims surrounding it. The first movement of Haydn's Symphony No. 45 in F# Minor ("Farewell") features a much noted, largely closed D-major episode in the last two-thirds of its development (Ex. 6). Whatever its motivic connections might be to earlier music, this self-contained episode, "important" or "highly etched" by any standard, provides an unexpected, placid contrast to the dogged "Sturm und Drang" texture that has prevailed up to that point. Were Cone's version of the sonata principle operative, this D-major episode should return "before the end of the movement" with the effect of tonic-oriented resolution (as in the *Eroica*). But this does not happen: once the development is over, the "interlude" theme is not sounded anywhere else in the movement.

Case closed? Not quite: in the case of the "Farewell" Symphony, no current discussion can stop here. In a well-known treatment of the piece, James Webster argued that the developmental "interlude" is one of the several features of instability and nonclosure in the first movement. He then proceeded to offer an elegant multimovement argument that the "interlude's aesthetic character prefigures the [tonic-major] 'farewells' [in the finale]," thus playing an important role in an impressively "through-composed" symphony. So far, so good. But then Webster took a second step: "Though not even hinting at a thematic recall, the [F#-major] ending [of the finale] thus implicitly *recapitulates* the interlude; the 'sonata principle' is affirmed, over the course of the entire symphony."²⁸ One supposes that this claim proceeds from an underly-

28. Webster, *Haydn's "Farewell" Symphony*, 39–45, 73–112, at 112. On the other hand, Haimo (*Haydn's Symphonic Forms*, 103, including n. 6) accounted for the finale's seeming look back at features of the earlier episode in terms of what he calls the unity principle, not the sonata principle. According to this interpretation, since Haydn rarely broke the motivic cohesiveness or unity of a sonata-form movement by introducing "dramatically different material" in the development, when he did so here it created an "extraordinary, anomalous event that could not be resolved within the movement itself." These lingering issues were eventually addressed at the end of the finale.

Example 6 Haydn, Symphony No. 45 in F# Minor (“Farewell”), first movement, mm. 108–15

ing assumption that if affirming the sonata principle over a single movement is a good thing, affirming it so cleverly over several movements must be an even more impressive compositional feat. More fundamentally, the assertion is also grounded in Cone’s shaky proposal that developmental episodes in that period called for tonal resolution (or, from another perspective, that when they *are* “resolved” they are responding to a bedrock demand for tonal balance within the style). If that proposal cannot be reasonably sustained—as I maintain in this portion of the present essay that it cannot—then this particular “Farewell” argument is left unsupported.

One understands Webster’s more general point, of course, and it is certainly not my intention to dispute that the conclusion of the “Farewell” is expressively related to the first-movement episode: one should be sympathetic to the desire to hear revisitings, variants, and cross-references over several movements. My concern, instead, is with the implications of a familiar style of language and the assumptions that appear to make the choice of that language possible. What does it mean, for instance, to “recapitulate” a developmental episode in a musical space far beyond the normative recapitulation (or, more problematically, to recapitulate it only “implicitly”)? In what sense is the formally specific, sonata-form term *recapitulates* the best choice of language here—especially in the concluding portion of this finale?

But complex cases make for unsound generalizations, and we may put aside these abstract speculations to return to simpler matters. Although Haydn’s “Farewell” Symphony is extraordinary in many respects, the existence of thematically episodic passages within developments is not. Such episodes are common, for example, in Mozart. This is especially true of, but not limited to, his early works, in which the development often begins with an episodic departure—more or less—from expositional material. The first movement of the String Quartet in G, K. 156, has been fairly widely cited and discussed in this regard.²⁹ Another relatively early instance (of many) may be found at the

29. See, for example, Alfred Einstein, *Mozart: His Character, His Work* (Oxford: Oxford University Press, 1945), 174; Hans Keller, “The Chamber Music,” in *The Mozart Companion*, ed. H. C. Robbins Landon and Donald Mitchell (New York: Norton, 1956), 90–137 (esp. 95–98);

onset of the developmental space of Mozart's Flute Quartet in C, K. Anh. 171 (285b): a tonally closed, eight-measure "new" sentence (syncopated rhythm, imitation in the bass) is stated in G minor (mm. 67–74), restated on D minor (mm. 75–83), and so on; this idea is never restated once the recapitulation begins (m. 111). A later, more famous instance is the "new theme" that appears at the opening of the development—and nowhere else—in the "Hunt" Quartet in B \flat , K. 458 (Ex. 7).

To be sure, upon reflection these episodes can sometimes be derived in one way or another from one or more of the expositional themes or underlying motives (as can both the "Farewell" Symphony interlude-theme and the new *Eroica* theme—determining the derivation of the latter has been standard fare for analysts over the decades). Finding such "latent relations," as Hans Keller called them, between any two themes conceived within the style—built from scalar fragments, triadic outlines, neighbor-note motions, and so on—is rarely a daunting task.³⁰ Normally, however, they are not mere audible variants of the expositionally established melodies; instead, their point is to impress us as something fresh—the onset of a new turn of events. (Hence the standard term for them: *episodes*.) Examples are easy to find, although different analysts might disagree with regard to either the absolute importance of the themes (are not all thematic occurrences important?) or the degree to which their singularity—and hence their need for resolution later in the work—might be explained away by the appeal to "latent relations."³¹

On infrequent occasions, when relatively new material opens the development, it may recur in the tonic in a coda. Consider, for instance, the first movement of Mozart's Piano Sonata in C, K. 330. Here the development begins with what seems to be a new idea in G major (mm. 59–64; Ex. 8), and the idea returns to provide the material for the brief C-major coda in the concluding bars, measures 145–50. One might suppose that in this situation, as in the *Eroica*, it is a confirming instance of Cone's sonata principle. (Charles

and Wolf-Dieter Seiffert, *Mozarts frühe Streichquartette* (Munich: Fink, 1992), e.g., 33–34. Cf. the more general remarks concerning developmental episodes in David Bushler, "Harmonic Structure in Mozart's Sonata-Form Developments," in *Mozart-Jahrbuch 1984/85* (Kassel: Bärenreiter, 1986), 18; and Charles Rosen, *Sonata Forms*, rev. ed., 275 ("There are many examples of new material introduced in the development section; Beethoven's *Eroica* is the most famous example, although here the new material is related to the main theme. We must distinguish . . . between development sections which introduce a characteristic and memorable new theme into the thematic development of material drawn from the exposition, and those which make no allusion to the exposition whatever").

30. Keller, "The Chamber Music," 91, 97. Cf. Rosen in note 31 below.

31. Cf. Charles Rosen, *Sonata Forms*, rev. ed., 288: "On the other hand, when the development contains new material, it, too, may be resolved in the recapitulation. This does not apply to new material immediately and obviously related to one of the principal themes of the exposition, as in Mozart's Sonata for Piano in F major, K. 332, but to material of a character evidently very different from anything in the exposition." (This passage did not appear in the original, 1980 edition of the book.)

Example 7 Mozart, String Quartet in B \flat Major, K. 458 (“Hunt”), first movement, mm. 91–98

Example 8 Mozart, Piano Sonata in C Major, K. 330, first movement, mm. 59–64

Rosen cited the passage as a “brief coda” whose “purpose” was “the recapitulation at the tonic of a new theme introduced in the development section . . . [satisfying] the demands both of symmetry and harmonic resolution.”³²

But here and in similar cases there is a more cogent explanation to be found. In this period, when a thematic coda was provided after the close of a rhetorical recapitulation (the layout of musical ideas parallel with that of the exposition), the most common option was to return to a tonic-key reference to the opening theme, as if beginning yet another ordered cycle through the basic materials. (More generally, this sense of immediately beginning another thematic cycle after reaching the final element of the layout—what I call the idea of *rotation*—plays a large background role within the style.)³³ In addition

32. *Ibid.*, 321–22.

33. The concept of rotation, for instance, also includes the (by no means invariable) tendency of material at the end of the exposition (or recapitulation) to return to the opening idea of the exposition, as if beginning a new thematic cycle (as 11:59 yielding to 12:00 begins a new rotation around the clock face). This is why both developments and codas often (but not always) begin with references to primary-theme material. The principle of rotation—along with instances where the principle does not apply—is discussed in Hepokoski and Darcy, *Elements of Sonata Theory*.

to this, considered in generic terms, such a primary-theme statement in the coda could also suggest a return to the idea that had begun the development section, since the most common option was to begin the development (rotationally) with an off-tonic presentation of primary-theme material. When this occurred, the onset of the coda could simultaneously refer to the practice of the development-recapitulation repeat. Indeed, codas that begin similarly to the openings of their preceding developments, while not mandatory, are frequently encountered.

Consequently, in cases in which the more typical primary-theme opening of the development has been written over with new, nontonic material, as happens in the first movement of K. 330, it is not surprising to find a coda launched with a tonic back-reference to that new theme. (One interesting feature of this movement is that the repeat of the second half follows the full statement of the coda. In itself this is not deviational.³⁴ In this case, though, it produces the result of sounding the coda-presentation of the “new” theme in the tonic, then recycling back to its off-tonic presentation for the repeat.) The main point is that in such situations the tonic arrival in the coda of the originally nontonic idea need not be taken as illustrating any presumed “demand” regarding the “recapitulatory” tonal resolution of developmental themes. Rather, that seeming resolution is better understood as a convenient by-product of a larger governing idea: that of thematic rotation, or the architectural propensity within the style to recycle arrays of thematic material in relatively the same order.³⁵

34. As Rosen correctly points out (*Sonata Forms*, rev. ed., 297), “There are essentially two kinds of codas: those that appear as a completely separate postscript to the second part after this has been repeated; and, somewhat less frequent, those that are placed at the end of the second part, but within the indications for repetition.”

35. Cf. the similar but more subtle case, also pointed out by Charles Rosen (*Sonata Forms*, rev. ed. only, 288 [cf. n. 31 above]), of the first movement of Mozart’s Sonata for Two Pianos, K. 448 (375a): “The development . . . opens with a new theme at the dominant and it is accordingly resolved at the end of the recapitulation.” Rosen wished to cite this instance as an illustration of the pervasive influence of his version of the sonata principle—a version to be discussed in the next section below—although in the immediate context (as already cited in n. 31 above) he worded his formulation of what was being illustrated cautiously: “When the development contains new material, it, too *may be* resolved in the recapitulation” (italics mine).

Even so, K. 448/i may be understood in more productive terms. Assuming that the situation mentioned above regarding K. 330/i more clearly articulates a norm—that is, a coda recalling the onset of the development—then the “developmental” music inserted following the close of the secondary theme in the recapitulation of K. 448 would be a typical example of the common practice of what may be termed “coda-rhetoric interpolation” (CRI): the early inclusion of typically codalike material before the appearance of the final thematic module(s) of the recapitulation proper. Such a device permits the exposition and recapitulation to end similarly while also including codalike material. (This happens, for example, at the end of the first movement of Mozart’s Symphony No. 40 in G Minor, K. 550.) A passage of CRI, therefore, is not most profitably understood as belonging to the recapitulation proper: it is in every respect an interpolation. (The principle of the CRI is discussed in Hepokoski and Darcy, *Elements of Sonata Theory*.) Cf. note 33 above on rotation.

Related considerations are also in play in the passage that Cone cited as his first example of the sonata principle, to which we may now return. His point, one presumes, was that in the coda to the first movement of the E \flat -major *Eroica*, a “new” and dramatically staged developmental theme originally stated in E minor (mm. 284–88, then modulating away from that key in a repetition) returns in E \flat minor (mm. 589–93, following an earlier sounding of it in F minor, mm. 581–85), seemingly resolving the earlier off-tonic appearance. Here one should observe that, as a whole, the *Eroica* coda retraces, in order, some notable moments of both the prior development and the recapitulation. (Reflecting this, it also features a modulatory first portion, ending with a long dominant preparation, and a stable, tonic second portion, beginning at measure 631, standing in for certain aspects of the recapitulation.) As such—besides accomplishing much more as well—it blends tonal and rhetorical features typical of discursive (lengthy, multisectional) codas with a thorough recasting and telescoping of the generically conceivable development-recapitulation repeat, a historically available, though increasingly rarely selected, option that Beethoven did not otherwise employ here.³⁶ Situated in the “development-reference” part of the *Eroica* coda, the “new” theme’s successive tracking through both F minor (mm. 581–85) and E \flat minor (mm. 589–93, the tonic minor, the central moment under discussion here) may be interpreted as a back-reference to the development’s triple-sounding of that theme first in two nontonic keys, then in the tonic minor: E minor (mm. 284–88), A minor (mm. 292–96), and E \flat minor (mm. 322–26). (The last of these follows a touching of the tonic, E \flat major, at measure 316 and a quick collapse into the parallel minor at measure 320. The whole passage, which leads to other harmonic shifts, eventually finds its way to the reactivation of the structural V at measure 338.) In sum, the E \flat -minor appearance of the “new” theme in the coda does not so much resolve that theme as reanimate its tonic-minor appearance earlier in the development.

Rather than citing additional examples, we might proceed to our first conclusions. *Pace* Cone, there is no reason to suppose that nontonic “new material” in a development necessitated any later tonal recapturing and resolution. Such isolated nontonic episodes were common within the style as impulsive digressions or reactions to past events, as compensations for expected material lacking in the exposition, or as instances of Tovey’s suggestion of the occasional need for relief within a composition. When this material did recur in the tonic in the coda, that appearance is better understood as responding to

36. This feature of the *Eroica* is discussed further (with additional examples from other works) in Hepokoski and Darcy, *Elements of Sonata Theory*, chap. 13 (“Paragenetic Spaces: Coda and Introduction”). A somewhat similar observation to this and what follows, though in a different interpretive context, was provided by Robert P. Morgan, “Coda as Culmination: The First Movement of the ‘Eroica’ Symphony,” in *Music Theory and the Exploration of the Past*, ed. Christopher Hatch and David W. Bernstein (Chicago: University of Chicago Press, 1993), esp. 371–72.

a different norm altogether, such as (but not limited to) that of rotation or the reference to the generic practice of block-repetition. Nor should one overlook the obvious: since the generic function of codas, when they existed at all, was to demonstrate the security of the tonic key finally fully attained in the preceding sonata form—even when, as often in Beethoven, they began with a temporary shift toward another key—one would expect that longer codas that reflect on events in the preceding structure would accomplish many of those backward glances in the tonic.

The larger point, though, is that it is in the nature of coda-spaces, when they refer (for any number of reasons) to past sonata-form events, to do so in the home key. This reference need not be construed as a response to any imagined imperative from the sonata principle. (Should examples turn up of clearly developmental “new” material being interpolated into a freshly created, separate modular space within the subsequent recapitulation—as opposed to a coda—those cases ought to be understood as exceptions [sonata deformations] that invite individual consideration.)³⁷ However we might construe the sonata principle, it should apply only to nontonic, post-medial-caesura zones of expositions and recapitulations, not to developments or (in virtually all cases) codas.³⁸ What may happen within codas—as striking artistic surplus, extravagant exception, or high workmanship—should not be taken as evidence for what must happen in them. But if it is a question of “may” rather than “must,” then we are not dealing with a fundamental principle but rather with an elegant add-on in certain highly crafted pieces.

This brings us to Cone’s more dubious proposal that the sonata principle can occasionally serve as a guideline governing the coda-space treatment of nontonic episodes in rondos. Here it is important to observe that Cone’s formulation of regulative stringency softened for these cases, away from an apparently underlying requirement for sonata forms to a merely “suggest[ed]” explanation of other individual cases. (Illustrations within rondos only occur

37. Cf. Rosen’s claim concerning the first movement of Mozart’s Piano Concerto in A Major, K. 488 (*Sonata Forms*, rev. ed. only, 288): “the new theme at the dominant that begins the development . . . also reappears in the recapitulation at the tonic.” But where the “development” begins in this concerto movement is anything but certain: in the recapitulation Mozart might have been playing on this ambiguity (hardly atypical in his concertos) by including this “*tutti* afterthought,” which had indeed gone on to dominate much of the development. In any event, such an unusual occurrence should not be adduced as exemplifying a norm, and we should recall Rosen’s caution in articulating this principle. (See n. 35 above.)

38. That the principle applies only to post-medial-caesura material is important. No matter what tonal suggestions it might make, transitional material (TR) of any sort, leading the primary theme to the medial caesura and secondary theme (S), was excluded from this guideline. (Some reasons for this differing status of TR- and S-ideas are also suggested in the third subsection of this essay, “Incomplete Secondary Themes in the Recapitulation: Two Axioms.”) This appears to be why, for instance, the nontonic theme from Beethoven’s Sonata in D, Op. 10 no. 3/i, mentioned earlier (n. 8) as something that precedes the functionally operative “S” proper, is exempted from the expectation of tonal resolution.

“sometimes.”) Now, as just proposed, one need proceed no further: what “may” happen within tonic-grounded codas (“sometimes”) should not be mistaken for a demand imposed on that coda by the preceding architectonic form. Yet in the context of Cone’s surrounding claims, his readers were being encouraged to understand the illustration at hand, taken from Mozart’s Piano Concerto K. 482, as an instance of the larger reach of the sonata principle, which “covers many aspects of formal treatment.” Amidst such assertions, it is clarifying to think more precisely about some basic features of the various types of rondos—and their potential codas.

Now obviously the “expositional” version of the sonata principle would be applicable, full strength, to the corresponding portions of sonata rondos. Here we should be clear about terminology. “Sonata rondo” is a description most appropriately applied to rondo-based compositions that have a clear exposition and recapitulation. Under this definition, sonata rondos are marked by an initial section that unfolds in the manner of a sonata exposition. Following a characteristic (and characteristically shaped) rondo theme, serving as the primary theme (P), we encounter a transition (TR) driving to a medial caesura (MC) and a secondary theme (S)—and sometimes one or more closing ideas (C) as well—before dissolving into a retransition that cycles back to the rondo theme. Thus its entire structure is: (1) exposition (beginning with P); (2) tonic-P reference and central episode or development; (3) recapitulation; and (4) tonic-P reference and coda (if any). In these commonly encountered structures the S- and C-material of the exposition is expected to be brought back, *in toto*, in the tonic in the recapitulation. (This is also true of the B section in the simpler, transitionless [and retransitionless] alternative—and in part, progenitor—of the sonata rondo, the “symmetrical seven-part rondo,” ABACAB’A-coda.)³⁹ We shall presume that this application of the sonata principle to sonata rondos is self-evident.

At stake in Cone’s proposal, though, are nontonic episodes in rondo-structure spaces that are not simultaneously parts of sonata expositions. This general category typically encompasses only two situations: central thematic episodes within sonata rondos (in the cases where that alternative is found, as opposed to a genuine development); and the B and (less often) C zones, if

39. The definitions in this paragraph follow those proposed in Hepokoski and Darcy, *Elements of Sonata Theory*. In that work we also refer to the true sonata rondo—with full exposition—as the “Type 4 sonata” and urge the employment of expositional labels for the zones and themes at hand. Objections to the more common—though unhelpfully reductive—descriptions of the sonata rondo itself as ABACAB’A have also been voiced elsewhere, perhaps most recently in Leon Plantinga, *Beethoven’s Concertos: History, Style, Performance* (New York: Norton, 1999), 82–84. For an example of the simpler ABACAB’A structure (not a true sonata rondo, though clearly moving in that direction), see, for example, the finale of Beethoven’s String Quartet in C Minor, Op. 18 no. 4. When the individual sections are even shorter, less expansive—as sometimes in earlier works—the piece is better regarded as a type of rondeau: for example, the ABACAB’A of the finale of Mozart’s String Quartet in C, K. 157, is probably best described as a “symmetrical three-couplet rondeau” (in which B’ is a tonal resolution of the nontonic B).

either is nontonic, in the five-part rondo, ABACA-coda. (In a five-part rondo, the C-episode not infrequently occurs in the tonic, though perhaps in the parallel major or minor; still, nontonic C-spaces—and tonic B-spaces—are possible.) Cone's suggestion was that these types of nontonic episodes were "sometimes" placed under the pressure of the sonata principle to reappear in the coda—or, stated more cautiously (to reflect his more modest verb, "suggests"), that when they do, that coda-appearance is to be ascribed to the influence of the sonata principle, the "underlying" postulate of the "Classical style," one that "covers many aspects of formal treatment."

Cone's claim was too accommodating. If the episode-theme or a snippet thereof returns in the coda, the sonata principle is declared to be confirmed; but if the theme does not return, the primacy of the principle within rondos is not to be regarded as falsified—indeed, its broad claims are not challenged in any appreciable way. It would seem that the principle is to be trotted out as the underpinning idea only when we find confirming cases.

But who would seriously claim that it was built into the expectations of sonata rondos and ABACA rondos that, upon hearing such a nontonic episode, we would anticipate that all or some of it would need to come around again, in the tonic, in a coda? Such an expectation is nowhere in sight in the sonata-rondo finales of Beethoven's first three piano concertos or the Violin Concerto: in none of these is a nontonic central episode revisited in the coda. Similarly, the G-minor central episode of the sonata-rondo finale of Mozart's Piano Sonata in B \flat , K. 281, never resurfaces later in the movement; nor does the broad, F-major central episode in the finale of Beethoven's Piano Sonata in C, Op. 2 no. 3. Within five-part rondos, the several minor-mode "Turkish" themes in the nontonic episodes of the finale of Mozart's Serenade in B \flat for Thirteen Winds, K. 361, are not referenced in the broad coda. Surely the reader can supply dozens of such instances, all of which would demonstrate that the sonata principle is not a central factor governing the behavior of rondo codas.

And yet, especially within larger codas, one does occasionally find tonic references to material from nontonic episodes. Such a reference may occur in differing gradations of completeness. It may be the citation of an entire thematic module (as in Cone's cited case of the "cadential tag" [better described in all of its contexts as a "*postcadential tag*" from B] in the coda of the ABA'CA' slow movement of Mozart's Piano Concerto in E \flat , K. 482); a briefer thematic reference folded into the general discourse (as in the five-part-rondo slow movement of Mozart's Piano Sonata in B \flat , K. 570); a transformation of only a characteristic motive from the earlier episode into clearly cadential material (as in the sonata-rondo finale of Beethoven's Piano Sonata in E \flat , Op. 7);⁴⁰ or, perhaps most subtly, the adaptation of only an accompaniment pattern from the nontonic episode to underpin a cadential module based on rondo-theme

40. Concerning Op. 7, see note 42 below.

material (as, perhaps, in the first half of the coda of Mozart's Rondo in A Minor, K. 511). Back-references come in varying degrees of explicitness and cadential completion, and any thoughtful analysis must distinguish among these degrees and their local implications.

What guides these compositional decisions? Not the sonata principle, but rather the enticing option, within certain codas in ambitious compositions, for larger codas to become retrospective—to reflect expressively upon that which has occurred in the preceding formal structure through a lens of valediction. (When the tonic key is minor, as in the C-minor slow movement of K. 482, the revisiting of a nontonic major-mode episodic idea in minor is particularly poignant—as if fatalistically mourning the loss of its once-major-mode sense of hope or confidence.) Alternatively, some larger Allegro codas might seek to be, at least in part, summary-codas, energetically drawing together into a final, surplus zone much of the diverse material of the now-completed structure. Whatever the case, as argued earlier, since codas exist primarily to ground the tonic—even though, exceptionally, there may be a temporary, local sidetrack elsewhere—one would expect that such back-glances would normally occur in the tonic key. What one encounters, therefore, is not so much a needed resolution of a nontonic theme as it is a free tonic recollection of something memorable that had occurred in the piece's past. Whatever sonata-principle impression it might provide is secondary—a side effect or surplus benefit of a more fundamental idea.

One last element within Cone's description of the sonata principle merits examination. This is his proposition that the presumed resolution near the end of a piece need not be a tonic resolution at all. It will suffice if the out-of-tonic "important statement" is merely "brought into a closer relation with the tonic, before the movement ends." Although at first glance this idea seems both engaging and appropriately flexible, it is difficult to determine the scope of what it is intended to encompass. What is meant by "closer relation"? How close is close? How would one go about explaining, for example, that the *Egmont* recapitulation's extended D \flat (implicitly "five-flat") sounding of the secondary and closing themes is tonally closer to the F-minor tonic than the exposition's A \flat ("four-flat") presentation of the same themes?⁴¹

Cone chose to illustrate the point, though, not with tonally unusual recapitulations but with the final bars of the finale of Beethoven's *Pathétique* Sonata. Until this moment of his discussion, the point of the sonata principle had been clear. It was to require certain nontonic "important statements"—presumably anchored with a theme, an episode, or a melodic "tag"—to recur in (or "closer" to) the tonic toward the end of a movement. But here he extended that idea to suggest the reverse. The principle was also to be considered operative if "before the movement ends" an originally *tonic*

41. Several issues surrounding this mediant-submediant relation in *Egmont* and other pieces—including that of the problematic significance of "fifth-transposition"—are discussed in Hepokoski, "Back and Forth from *Egmont*" (n. 10 above).

theme, such as a primary theme or a rondo theme, touched on a preceding, memorably enunciated *nontonic* area. This meant that the sonata principle was to allow a set of reciprocal possibilities, which Cone apparently regarded as complementary, not contradictory. Not only were “important” nontonic ideas expected at some later point to become tonic ones (or ones sounded in “close” tonal substitutes), but an early tonic idea could relieve a mid-movement, nontonic theme from the presumed expectation of resolution if the former were sounded, however briefly, in the relevant nontonic key. Thus, at the close of the *Pathétique* finale (a sonata rondo in C minor with an interior episode beginning in A♭ major, mm. 79ff.), we find a smoothly integrated reference (mm. 203–6) to the nontonic A♭ major, marked with a variant of the head-motive of the refrain, as Cone observed, not with the original episodic theme (Ex. 9).

To be sure, this A♭ reference in the coda is suggestive of much that has preceded it in ways that may be readily elaborated. Nearly everyone could agree, I suspect, that this fleeting A♭ module, strikingly prepared by its own dominant seventh, looks back at the key of the earlier episode. (More than that, it probably also recalls the key of the second movement, the A♭ Adagio cantabile.) Do we need to make further claims for this passage? Would we consider this conclusion—and the movement—to be in any way wanting (or at least unresponsive to a basic “principle”) had Beethoven not provided this submediant allusion? Did he really call forth this patch of A♭ color to invoke or respond to a background tonal postulate lying at the center of his compositional practice? If so, then we would expect to find, without much difficulty, such primary-theme, nontonic gestures toward the key of interior episodes occurring as more or less standard practice in codas to sonata, rondo, or sonata-rondo structures. And this, surely, is not the case.

One might be more inclined to understand the *Pathétique* example as a compressed and richly allusive instance of early Beethoven’s occasional fondness for brief “wrong-key” statements of portions of the main theme toward the ends of sonata-rondo finales. In the fourth movement of the Piano Sonata in C, Op. 2 no. 3 (with interior episode in the subdominant, F major, mm. 103ff.), for example, one finds, following a fermata in measure 297, the rondo theme sounded not in the interior episode’s F major but rather on A major, VI, decaying to A minor, vi, before being brusquely integrated back into C major for the concluding cadence. In the fourth movement of the Sonata in E♭, Op. 7 (with interior episode in C minor, mm. 64–88), the reprise of the last rounded-binary presentation of the rondo theme is first heard, tentatively, on E major (= ♭II, mm. 155–60, beginning over the dominant) before settling back to E♭. (In this case, as mentioned earlier, portions of the interior episode return in the tonic shortly thereafter, in measures 166–83.)⁴²

42. My contention here, obviously, is that any interpretation of the ephemeral A♭ moment at the end of Op. 13/iii should be informed by an understanding of such possibly related precedents as we find in the final pages of Op. 2 no. 3/iv and Op. 7/iv. Of these two movements, the former may need no additional commentary here. With its final return to material from the C-minor

Example 9 Beethoven, Piano Sonata in C Minor, Op. 13 (*Pathétique*), finale, mm. 200–210

The image shows a musical score for the finale of Beethoven's Piano Sonata in C Minor, Op. 13 (Pathétique), measures 200–210. The score is in C minor and 3/4 time. It features a piano part with dynamic markings *sf*, *p*, and *decresc.*, and a bass part with dynamic markings *pp* and *ff*. The score includes fingerings (6, 7, 3) and articulation marks.

On the other hand, in this matter of the presumed “closer-relation” possibility, I do not wish to overreach in my reading of Cone, who goes out of his way to mention that Beethoven only “sometimes” does this and to describe the illustration from the *Pathétique* as a “special bow.” Implicitly embedded in

interior episode in the $E\flat$ tonic (mm. 166–83), however, Op. 7/iv may be more instructive in an attempt to propose that the apparent sonata-principle effect at hand is instead a satisfying by-product of a more fundamental procedure—that of rotation.

In the Op. 7/iv sonata rondo we find four rondo “cycles” (rotations), each beginning with the rondo theme: the first at measure 1 (producing an exposition); the second at measure 51 (incomplete rondo theme leading to an interior episode in C minor); the third at measure 94 (producing a recapitulation); and the fourth at measure 143 (final statement, varied, leading to a coda-close based on the interior episode). This concept of four rondo-rotations is resonant with early nineteenth-century discussions of the sonata rondo as found, for instance, in Reicha and Czerny. (See, e.g., the discussions in Malcolm S. Cole, “Sonata-Rondo, the Formulation of a Theoretical Concept in the Eighteenth and Nineteenth Centuries,” *The Musical Quarterly* 55 [1969]: 180–92 [esp. Reicha’s groupings in the “Coupe du Rondeau”]; and idem, “Czerny’s Illustrated Description of the Rondo or Finale,” *The Music Review* 36 [1975]: 5–16 [esp. the groupings in the “model-rondo” diagram, p. 5]). But here the expositional first rotation (ordered layout of materials) corresponds closely with the recapitulatory third, and the second, as it happens, corresponds with the fourth. In the second, the rounded-binary theme (m. 51) is aborted before the reprise with a half-step shift from the dominant $B\flat$ (m. 62) to $B\sharp$ (m. 63), which thereupon ushers in the C-minor episode (m. 64). In the fourth, we find an expansion of the same succession of materials: the onset of the rondo theme (m. 143); the half-step shift on the dominant, $B\flat$ to $B\sharp$, at the same structural location (mm. 154–55), although this time the reprise—the last limb of the rondo theme—is sounded, at first on E (largely over its dominant), then shifting down to $E\flat$ (mm. 161–62) to produce the closing cadence in measure 166; and the figural elements from the interior episode, now in the tonic (mm. 166–83). In other words, the larger organizational basis for the movement’s large-scale layout is a broader double-rotation of two rondo cycles each: 1–2 and 3–4. The tonic return of the interior-episode figuration at the end makes this compound-rotational aspect—always potentially present in sonata rondos with discursive codas—more explicit than one finds in other pieces. The potential for rotational implications within rondos is dealt with in greater length in Hepokoski and Darcy, *Elements of Sonata Theory*.

this wording, though, is that when such occurrences do happen, they are to be taken as suggestions of this adaptation of the sonata principle; when they do not—as is more normally the case—the principle need not be considered as challenged and the contrary inference is not to be drawn. This is not what one would expect of a “unifying [and] underlying . . . principle for the Classical style.” Cone’s “closer-relation” proposal—at least with regard to the *Pathétique* finale—remains underdeveloped and, it would seem, unpersuasive.

Whatever the sonata principle might be, within music from the decades around 1800 it is properly applied only to expectations normatively placed on expositional nontonic themes following the medial caesura (that is, from the “secondary theme” onward).⁴³ Nontonic new themes or “important statements” within developments or interior episodes are required by no principle to return later in the movement (much less, later in the composition). When they do, it is that return that is exceptional and must be accounted for. Under no circumstances should that return be cited as the validating of a norm. The sonata principle is irrelevant to considerations of central episodes in sonata rondos or B and C sections within five-part (ABACA) rondos.

Rosen: Basic Texts and an Example

An idea similar to Cone’s, although not accompanied by the term *sonata principle*, was put into a clearer focus by Charles Rosen in his two much-read, neo-Toveyan studies from 1971–72 and 1980: *The Classical Style* (rev. 1997) and *Sonata Forms* (rev. 1988).⁴⁴ For the most part it has been Rosen’s version of the concept that has persisted within the discipline. As is well known, Rosen insisted that the tonic and dominant keys within sonata expositions were related in terms of conflict, as “a polarization or opposition” (instead of, say, as a complementary set or as a tonic moving outward to one of its natural ramifications, an altered projection-of-self—interpretations that I find preferable). Within an exposition this polarization of keys generated “a large-scale dissonance: the material played outside the tonic (i.e., in the second group) is dissonant with respect to the center of stability, or tonic.” What was proposed was the “concept of *dissonant section*” (Rosen’s italics)—a structural dissonance to be reconfronted and brought into the tonic later in the movement (normally in the recapitulation).⁴⁵ Here we have a new formulation of the sonata principle, wisely restricted only to “second groups” within expositions and unencumbered, for the most part, by veiled remarks concerning restatements in a “closer relation” to the tonic.

43. On the importance of the “MC” see Hepokoski and Darcy, “The Medial Caesura” (n. 10 above). On the post-medial-caesura aspect of the guideline, see note 38 above.

44. Even the pointed plural, “sonata forms,” for instance, recalls the title of Tovey’s *Encyclopaedia Britannica* articles. Cf. note 12 above.

45. Rosen, *Sonata Forms*, rev. ed., 229. Cf. p. 287.

In *Sonata Forms* the principle, with some flexibility, was laid out as follows:

Sonata style by the 1750s turns this modulation [from tonic to dominant] into an overt confrontation of tonalities: that is, the area in an exposition governed by the tonic is firmly distinguished from that governed by the dominant (even when, as often in Haydn, the transition from one to the other may be very extensive), *and all the material played in the dominant is consequently conceived as dissonant, i.e., requiring resolution by a later transposition to the tonic.* (Rosen's italics.)

and

What must reappear in the recapitulation—and this is a rule that holds true from the very beginnings of anything that can be called sonata style—is the second group, at least any part of it that has an individual and characteristic aspect, and that does not already have its analogue in the first group. The resolution of this material confirms the articulation of the exposition into stable and dissonant sections. A theme that has been played only at the dominant is a structural dissonance, unresolved until it has been transposed to the tonic.⁴⁶

Rosen's wording of the "rule" ("less a compositional canon . . . than a sense of the aesthetic balance essential to late eighteenth-century style")⁴⁷ was certainly clear, and as an almost self-evident proposition it covers the majority of sonata expositions and recapitulations. The problem was that the principle was violated by a small number of prominent compositions, especially (but not only) ones by the unpredictable Haydn, whose often-recomposed recapitulations elude capture by such generalizations. And yet, since the rule was supposedly fundamental to the "sonata style," and since it also commended itself by sidestepping the trap of considering sonata forms as fixed schemata, it was not to be lightly swept aside. Rosen was faced with a dilemma. On the one hand, he could soften the claims of the principle to make room for legitimate exceptions. This, in some wordings, he implicitly did (although to admit that the principle could be eclipsed by a different consideration, as Haimo would later argue, was to deny it primary status). On the other hand, he could devise ways of interpreting the exceptions as still paying homage to the idea, however it might have to be recast. This was Rosen's more typical strategy. In both of his books he sought to justify a few exceptional cases and render them normative under the principle. It is here that his analyses fall short.

As a way into this issue, we might back up to *The Classical Style* of 1971–72, where the same concepts had been formulated:

It is the classical sense for large areas of stability, impossible before and lost since, that establishes what might seem to be the one fixed rule of sonata recapitulation: material originally exposed in the dominant must be represented in the tonic fairly completely, even if rewritten and reordered, and only material exposed in the tonic may be omitted. This is, of course, not a rule at all but a

46. *Ibid.*, 25, 287.

47. *Ibid.*, 288.

sensitivity to tonal relationships. . . . Material presented outside the tonic must have created, in the eighteenth century, a feeling of instability which demanded to be resolved. When the tonic was reaffirmed in the second half of the piece, the material already presented in the tonic could be, and often was, drastically cut, but the rest of the exposition cried out for resolution in the tonic. Today, our harmonic sensibilities have become coarsened by the tonal instability of music after the death of Beethoven, and the strength of this feeling is perhaps difficult to recapture.⁴⁸

Here the words “fairly completely” propose a flexibility also suggested in the version given in the later *Sonata Forms*: presumably a characteristic portion of the secondary theme (henceforth, S) could be employed to stand for the whole—something on the order of a synecdochic strategy. This important proviso might have been pursued to engaging conclusions (it does seem to be correct), but Rosen did not follow it up. On the contrary, he sought to defend the stricter version of the concept, whereby we would expect everything within S to be accounted for in the recapitulation. Let us turn to Rosen’s wrestling with an exceptional instance that would appear to challenge the fundamental idea. What follows is a pivotal, and characteristic, extract from *The Classical Style*.

It is worth examining this in some detail, at least briefly. First, for an exception to prove a rule. There is one Haydn quartet, op. 64 no. 3 in B flat [first movement], in which one of the second subjects appears nowhere in the recapitulation. . . . The first melody in the dominant, F major, is also the first regular-sounding melody in the quartet (mm. 33–42). A four-measure phrase, it is played first in the major and immediately repeated in the minor, and it clearly functions in the exposition to reaffirm the dominant. (It is not the only theme so used: the opening theme is replayed in the new tonality, and yet another new theme is then introduced, also in F major.) The repeated four-measure phrase does not, as I said, reappear in the recapitulation, but it does, however, reappear *in its full form* in the development section, *and on the tonic*. This time the phrase is played twice in the minor. In this way the theme is satisfactorily recapitulated, as one half of it was already in the minor to begin with; in addition, the tonic major is avoided in the development. All the various classical demands for balance and tonal resolution have thus been reconciled.⁴⁹ (Rosen’s italics)

Haydn stages the exposition of the first movement of Op. 64 no. 3 in a typically ingenious manner. Following a medial-caesura effect (MC) in measure 32 (built around a half cadence in V; see Ex. 10), S¹ emerges in the dominant, F, as a swaying, *dolce*, *Ländler*-like melody, what Rosen called “the first regular-sounding melody in the quartet.” It begins with a four-bar basic idea (mm. 33–36) that oscillates between tonic and dominant. But with its attempted repetition in measure 37 (whatever the outcome of that repetition might have

48. Rosen, *The Classical Style: Haydn, Mozart, Beethoven* (New York: Norton, 1971–72; also rev. ed., 1997), 72–73.

49. *Ibid.*, 73.

Example 10 Haydn, String Quartet in B \flat Major, Op. 64 no. 3, first movement, mm. 30–57

30 *fz*

34

38

42

Example 10 continued

been), the F major decays instantly to F minor: Haydn thereby provides a sign that the guilelessness of S^1 has hit a snag (a “lights-out” modal effect).⁵⁰ By measure 40—seeking to twist free of F minor—the harmony of the now-distorted basic idea shifts onto a $D\flat$ chord (VI of F minor), which is subse-

50. It is possible to be more precise regarding terminology. If we were to follow the classifications of Caplin (*Classical Form*, 35–70), measures 33–36 are best not regarded as an antecedent (they do not end with a clear cadence as normally defined; indeed they end on a V_2^6 chord in F major, m. 36) but as a compound basic idea (cbi). What we have, adapting these terms, is proba-

quently converted into an augmented sixth within F (m. 41) that releases onto a second medial-caesura effect (MC) at the downbeat of measure 42 (again constructed around a half cadence in F), announcing the imminent arrival of a “new” secondary theme. This new S begins in measure 48, after six bars of prolonged caesura-fill over the dominant (mm. 42–47, based on the anacrusis of P, recalling events earlier in the movement—cf. mm. 23–27—and restoring the major mode).

The S¹ basic idea (mm. 33–36) and its varied, distorted repetition (mm. 37–42) represent not so much a “regular-sounding melody” as a basic idea destabilized by modal decay. Far from serving “to reaffirm the dominant [key],” as Rosen claimed, what it actually illustrates is the opposite, namely, the difficulty that *this* S¹ is having in its task of reaffirming F major in the expected way, that is, by proceeding to a major-mode perfect authentic cadence. S¹ may begin as blissful security, but the composer has it soon run into problems of insufficiency. One might interpret the music as suggesting that the decay to F minor in measures 37–42 represents a refusal to continue the perhaps overly serene S¹—a dissatisfied jettisoning of that idea and the beginning of a search for another. (To be sure, secondary themes that destabilize into the minor are not uncommon within the style as part of a strategy of extension. In each case, though, such a decay has telling local implications that we should pursue hermeneutically. It is not merely an expressively neutral option to be passed over casually.)

As I have demonstrated elsewhere, this scenario of an insufficient original theme is a purposeful by-product of expositions that contain “apparent double medial caesuras” (or vice versa: the second caesura could be the planned outcome of the thematic process).⁵¹ Here the two MCs occur in measures 32 and 42. More commonly, expositions contain only one MC, dividing the exposition into two clean halves, the second of which begins with the secondary theme, S. The presence of two apparent MCs complicates the situation. In every case the double-MC effect stages the rejection of the theme immediately after the first MC. (In this case, but not in all, that first theme functions as S¹—a satisfactorily prepared “secondary theme.” In other cases the more convincingly “real” S¹ occurs only after the second MC effect.) The rejection invariably consists of a dissolution of the now-problematic ongoing theme, the avoidance of an authentic cadence, and the move into a new, transitionlike section. Sometimes this reinvigorated TR-function, as here, suggests a backing

bly a *cbi* + dissolving restatement (mm. 33–36, 37–42). In a personal communication to me regarding this theme, Caplin also noted that “although the ‘repetition’ of the 4-m. basic idea fails . . . the melodic content of the whole idea is still largely preserved, though now presented with augmented note values.” As is probably also evident, throughout this essay, I am only slightly adapting terms coined by Caplin—*presentation*, *continuation*, and the like—to describe sentence structure in this and other works.

51. Hepokoski and Darcy, “The Medial Caesura,” 145–50 (the subsection “The Mid-Expositional Trimodular Block”). See also Hepokoski and Darcy, *Elements of Sonata Theory*.

up to recapture certain pre-first-MC features of the music. Most important, though, is that it also sets up a second medial caesura and, shortly thereafter, a new, more cadentially successful S-theme (or set of linked S-modules).

The apparent-double-MC effect produces what I have referred to as a tri-modular block (TMB). This is a familiar expositional complication in which the label TM^1 stands for the theme after the first apparent medial caesura or its near accomplishment (a theme that does not reach closure in a perfect authentic cadence), TM^2 represents its dissolution and “transitional” preparation for the new MC, and TM^3 marks the new, more successful S-idea after the second-caesura effect. In Op. 64 no. 3 the second part of the exposition is thus complicated by a TMB situation, one in which the first-proposed S is declared to be inadequate and is set aside in pursuit of a new one. The new secondary-theme idea in the dominant (m. 48, TM^3) suggests a fresh resolve, a vigorous propulsion toward the anticipated perfect authentic cadence in F major. Notwithstanding its energetic scrambling, TM^3 also recalls certain harmonic aspects of TM^1 (S^1). It begins, for instance, with four bars of tonic-dominant oscillation, a static “warm-up” (cf. also mm. 8–9) before a more melodic idea takes off in measure 52—the “shot arrow” of continuation now let loose in its trajectory toward the cadence. (As such, TM^3 marks the initiation of a new sentence, in which the presentation is “melodically empty,” in contrast with the material at measure 33.)⁵² The perfect authentic cadence bringing an end to secondary-theme space (the generic point of “essential expositional closure,” EEC) occurs in measure 65 (not shown in the example), and a short retransition based on the primary theme, P, follows.⁵³

Every TMB situation presents a challenge in the recapitulation, and each solution is tailored to the particulars of the piece at hand. In this instance Haydn recomposed the recapitulation in such a way as: (a) to suppress the first of the apparent medial caesuras altogether, thereby taking care not to produce a TMB here; (b) to omit, as a consequence, the problematic TM^1 (S^1) and its minor-mode decay entirely; (c) to recompose the TR-approach to what is now the only MC effect, occurring in measure 152; and (d) to omit from S-space the static “warm-up” of what had been the beginning of TM^3 in order to leap directly in measure 153 into a transposition of measure 52 (the “shot arrow” from TM^3), from which crux-point all subsequent measures, however varied, become referential to their predecessors in the exposition.⁵⁴ (Material from the excised measures at the opening of TM^3 is put to use earlier by transplanting their texture into the recomposed TR, measures 141–46). There is no

52. This observation is indebted to a personal communication from William E. Caplin, from whom I also borrow the phrase “melodically empty.”

53. On the concept of the EEC, see Hepokoski and Darcy, “The Medial Caesura,” and, especially, *Elements of Sonata Theory*; and Hepokoski, “Back and Forth from *Egmont*.”

54. This concept of “crux” is built on that found in Ralph Kirkpatrick, who used it to refer to the similar point within the second part of the keyboard sonatas of Domenico Scarlatti. See his *Domenico Scarlatti* (Princeton: Princeton University Press, 1953), 253–61.

coda proper: the original retransition is converted into a closing-theme, “codetta” phrase. In brief: in the recapitulation Haydn rewrote the transition, omitted S^1 (TM^1), and rejoined expositional material only at the original measure 52. Within S-space he suppressed the first nineteen bars of secondary-theme material (all of TM^1 and TM^2 ; the opening of TM^3) and retained only the last fourteen (varied). Obviously, this outcome violates both Cone’s and Rosen’s versions of the sonata principle.

Rather than modifying the claims of the principle, Rosen sought to legitimize the strong version of it by noticing the important point that the omitted S^1 had made a post-expositional appearance in the development—and in $B\flat$ minor, the tonic minor—hence, apparently, obviating the need for a redundant tonic appearance in the recapitulation. The passage in question begins some eighteen bars into the development, and it is shown in Example 11. Before proceeding, it will be useful to notice that the development as a whole is rotational, that is, it touches on expositional material, for the most part, in the order in which it was originally presented. (In order to be considered a rotational development, such an architectural space need not be, and usually is not, thematically complete but does need to contain representative material from both halves of the exposition.) Setting aside the consideration of key-areas touched upon, we have: P^1 (m. 70), P^2 (m. 78), S^1 (m. 87), P^2 again (m. 97, perhaps an exception to the strictly rotational presentation of themes, though one should also notice the “emerging-augmented-sixth” resemblance of mm. 97–104 to the setup and beginning of TM^2 , mm. 40–42), the TM^3 opening (m. 105, not the TM^3 “arrow”), and the retransition (m. 118, itself P -based). The recapitulation (and a new rotation) begins with P^1 in measure 126.

As Rosen noted, in measures 87–90 we hear a variant of S^1 on the tonic minor. Perhaps the first thing to observe, though, is the weak “tonic”-status of that $B\flat$ minor. Viewed only slightly more broadly, that key emerges here as only what Schenkerians might regard as an apparent tonic, since it is the second element of a larger I–ii–V–i progression in $A\flat$ major: measures 78, 87, 94, and 97.⁵⁵ In other words, this $B\flat$ minor is locally subordinated in clearly audible ways to a larger prolongation of a nontonic key. This alone problematizes any simple claim regarding a potential tonic statement of S^1 at this point. In addition, the theme’s MC-effect preparation is virtually nonexistent. This is an important feature, since it also suggests a weakened rhetorical launching of the S^1 -module.

Given the past history of this once-“innocent” theme—its having been marked for minor-mode decay and extinction in the subsequent expositional

55. Distinctions between real and apparent tonics within developments, for instance (although usually concerning a different situation from the present one: see n. 61 below), were outlined and illustrated in Jack Adrian, “The Ternary-Sonata Form,” *Journal of Music Theory* 34 (1991): 57–80; and idem, “The Function of the Apparent Tonic at the Beginning of Development Sections,” *Intégral* 5 (1991): 1–53.

Example 11 Haydn, String Quartet in B \flat Major, Op. 64 no. 3, first movement, mm. 86–107

The image displays a musical score for Haydn's String Quartet in B \flat Major, Op. 64 no. 3, first movement, measures 86–107. The score is arranged in four systems, each containing four staves (Violin I, Violin II, Viola, and Cello/Double Bass). The key signature is B \flat major (two flats) and the time signature is 3/4. The first system (measures 86-89) shows the beginning of a phrase with various rhythmic patterns. The second system (measures 90-93) continues the phrase with similar rhythmic motifs. The third system (measures 94-97) features a dynamic shift to *p* (piano) and includes a *crescendo* marking. The fourth system (measures 98-101) shows a dynamic shift to *f* (forte) and includes a *crescendo* marking. The score includes various musical notations such as slurs, ties, and dynamic markings.

Example 11 continued

module—the central thing to notice is that it now begins in minor. The theme has lost its capacity for major-mode statement, however evanescent. As a result Haydn has marked it as decayed, fallen from its original condition in the exposition. Moreover, in the exposition the minor-mode transformation had coincided with the process of discarding this theme in favor of another, major-mode one. Particularly when guided also with the foreknowledge of what is to come in the recapitulation, we might suppose that this more fully saturated minor-mode presentation in the development prepares for the possibility of the eventual liquidation of the theme altogether.

As in the exposition, the subsequent repetition of S^1 (beginning in measure 91) continues on $B\flat$ minor, and, modeled after the exposition, it seeks to wrest free of the minor mode through a quick shift toward a related major key via the “hopeful” $V^7/A\flat$ major in measure 94 (cf. mm. 40–42). But the push toward major is undermined: it is $A\flat$ *minor* that is sounded, chillingly *piano*, in measure 97 (more signs of dissolution), and through rising inner-voice chromatic motion that $A\flat$ (recalling the exposition) is driven to a strong V/C

minor by means of an unmistakable caesura-effect in measure 104.⁵⁶ Rosen's claim that the S^1 theme "reappear[s] *in its full form* in the development section" is generally true in terms of thematic contour—although that feature is not peculiar to this piece. On the other hand, as mentioned earlier, merely to claim, with Rosen, that the theme occurs "*on the tonic*" simplifies things unacceptably. It only begins on the apparent-tonic minor, then shifts toward $A\flat$ minor-major and eventually winds up articulating V/C minor—thereby producing another bleak minor mode looming ahead—as if the minor mode were now inescapable. S^1 and its consequences may begin on $B\flat$ minor, but the more important point is that this $B\flat$ minor, itself only an apparent tonic within the broader scheme of things, is not secured by any authentic cadence. Instead of $B\flat$ minor *tout court* what we have is a vision of that key, unsecured, that slips away elsewhere. No theme lacking an unequivocal authentic-cadential mooring should be regarded as resolved. In fact, the opposite is the case: not to bring the theme to such a cadence is to demonstrate how its potential for resolution has vanished. Rosen's assertion that this might-have-been tonic appearance signifies that "the theme is [thereby] satisfactorily recapitulated" misses the point. As a consequence his claim that "all the various classical demands for balance and tonal resolution have thus been reconciled" is overstated.

A more adequate discussion of this piece would dwell on the self-imposed problem of the trimodular block (TMB) in the exposition and its consequences. If one is prepared to hear the TMB complication as refractory-material-to-be-worked-with—something to be kneaded out in subsequent rotations—then many of Haydn's compositional decisions in Op. 64 no. 3 not only make more sense but may strike us as more creative. Assume, for instance, that it had been Haydn's plan from the outset to remodel his recapitulation in such a way as to smooth out the expositional TMB. In that case, part of the strategy of the piece might have been to demonstrate the progressive liquidation of the expositional element that was to be omitted, S^1 . Toward that end, Haydn made use of successive reinterpretations of the S^1 idea in the minor mode, as a sign of dissolution. One may observe the progression of S^1 through the piece's three thematic rotations: major-minor (exposition), minor-minor (development), and nonexistence (recapitulation).

The added feature of a noncadential *tonic* minor in the development could be understood as an enhancement to this ongoing S^1 effacement. It is also reasonable to argue (with Rosen) that its touching on the tonic (minor) here

56. Notice the replication of the shape, though not the content, of the exposition's trimodular block in the development. Here the double-MC effect is weakened by an unsatisfactory—perhaps nonexistent—caesura-effect at measures 86–87. What follows is tonic-minor TM^1 (S^1), TM^2 merging this time into a transitional variant of P^2 (m. 97), a notably strong MC-effect (m. 104, V/C minor), and TM^3 sounded out-of-tonic, in C minor (m. 105). This succession of in-tonic and out-of-tonic themes, along with modified caesura-practice, will be relevant to the larger argument concerning tonic S -statements proposed in the following section of this essay.

does help to compensate in advance for its absence in the recapitulation. But if one makes this claim offhandedly, as if its further implications were self-evident, one sidesteps a more robust confrontation with the larger issues of thematic placement and tonal resolution that are at the heart of the piece. In turn, this permits the reader to conclude that the assigned locations of tonic S-material result from arbitrary compositional decisions. To accept such a conclusion, however, would undercut the *raison d'être* of most analysis itself, which generally presupposes, at least heuristically, that compositional choices are made with purposes in mind that are worthwhile to seek to recover—if only as proposed reconstructions in the analyst's mind. The first movement of Op. 64 no. 3 actually participates in the larger compositional practice (though a rare one) of providing tonic S-presentations in the development, and it is the broader purpose of this compositional option and its structural implications that needs to be more carefully examined. In order to bring the current discussion of Op. 64 no. 3 to a temporary close, I shall defer that question to the next section, which will reengage this quartet movement along with other, complementary works mentioned by Rosen.

For the moment we may conclude that the upshot of Haydn's tinkering in Op. 64 no. 3 was to reformulate the initial expositional *Anlage* (thematic layout) to suggest a correction of the original double-medial-caesura/TMB situation, with its problematized S¹. The result was a cleverly recast recapitulatory space that is more normative, or at least less aberrant, within the style. This ameliorative concern, typical of Haydn's recapitulatory practice, suggests a background idea that is stronger than the sonata principle per se (which does not encourage inquiries into local thematic function or generic meaning). We may formulate it thus: One purpose of many normative recapitulations is to restore order or smooth out tonal or rhetorical problems set up in the exposition. (Recapitulations that intensify them further—as occasionally with much expanded TR-sections—are less frequently encountered.) The most obvious instance of this is the orthodox tonal resolution of the secondary and closing themes (the post-MC-space). This is a simple matter when the layout of the recapitulation generally matches that of the exposition. But when the layout is markedly different—when one finds reordered modules, omitted modules, suppressed or added medial-caesura effects, newly added sections of *Fortspinnung*, and so on—then a different game is afoot. Haydn often loaded an expositional *Anlage* with generic strains or surprises (medial-caesura problems, tonal or modal decays, frustrating cadential evasions, contrary-to-norm dynamic indications, and so on), and it would be the task of the development and especially the recapitulation to deal with them.

Haydn's recapitulations may often be productively viewed as ingenious readjustments of wittily problematized expositions. (Here one should remain wary of ideologically loaded talk of “perfect” resolutions and “fully satisfactory” closures of all expositional elements.) If that readjustment entailed the jettisoning of expositional material from S-space, so be it, so long as at least

part of S-space was retained—normally its concluding modules, whose task it is to drive to the crucial structural goal, the first satisfactory perfect authentic cadence (the point of essential expositional closure, the EEC). This, again, may be termed the synecdochic strategy, whereby a part of the original S is used to stand for the recovery and resolution of S-space.⁵⁷ This is not an idea to be invoked lightly. I am arguing neither that the synecdochic strategy is commonly encountered in recapitulatory S-space nor that its employment in any given piece is a casual matter. On the contrary, such deformational practice drives to the heart of what a given movement is musically and generically “about.” Incomplete S- or C-spaces within recapitulations are surely to be understood as non-normative. Unless we are prepared to assert that compositional practice, at bottom, is arbitrary, we must presume that each instance responds reactively to local conditions generated earlier in the work. (In Op. 64 no. 3 these are the implications embedded in the way that the TMB situation had been staged in the exposition.) Hence as an initial axiom of analysis one should suspect that any incomplete recapitulatory S- or C-space emerges under a larger narrative purpose or scheme of conceptual organization. It is our hermeneutic task to reconstruct this purpose.

The analytical trick is to be able to identify what the original expositional problems might be—the problems whose solution, full or partial, governs much of the rest of the movement. That identification demands a knowledge of the different types of standard expositional practice—along with Haydn’s individualized customization of those types—and their local norms and options. Apart from that, one might merely restate the obvious in reminding ourselves that, given Haydn’s obsession with surprise and originality, each movement is unique in its manner of confrontation with these issues. Every piece exists in dialogue with a constellation of generic norms in different, individualized ways. Traditional appeals to the quick solutions offered by the sonata principle will not do.

Incomplete Secondary Themes in the Recapitulation: Two Axioms

Op. 64 no. 3 was one of three exceptional cases mentioned by Rosen in *The Classical Style*. The second was the first movement of Haydn’s Quartet in D, Op. 50 no. 6, regarding which he wrote that “four measures of the exposition (26–29) are in the dominant minor, and again they are not in the recapitulation; again, however, they appear in the *tonic* minor in the development sec-

57. More generally construed, the synecdochic strategy would encompass all references to the presence of a whole (usually, a whole expositional section) through a partial articulation or representation of one or more of its parts. The strategy is often employed in rotational developments (not all of S and/or C need appear in them to be considered rotational), and it also applies, of course, to such things as shortened P-zones or P-TR mergers within recapitulations.

tion.” And the third was the first movement of the Quartet in G, Op. 77 no. 1, in which

the opening subject is repeated in the exposition at the dominant [as the secondary theme] with the theme in the cello; this is the way it appears in the tonic in the false reprise [in the development], and accordingly it does not have to be recapitulated in this form later. In this same movement there is a further example of the absence of a theme in the ‘second group’ from the recapitulation: again it is played in the tonic (major) in the development, but only at the end of the development, as it is used to re-establish the tonic and reintroduce the main theme.

In sum: “These are the rare cases in the Haydn quartets of material exposed in the dominant and missing from the recapitulation, and at each point we have seen that some form of tonic recapitulation has been provided.”⁵⁸ The assertion was nuanced—and yet concretized—further in *Sonata Forms*:

In the few cases where [this sense of “aesthetic balance”] appears to be broken, either the theme of the second group which does not appear in the recapitulation is replaced there by a passage of significantly similar harmonic character and shape (e.g., Haydn’s Symphony no. 75, where bars 52–55 are represented harmonically by 135–40), or else the theme has already appeared *at the tonic* in the development section, as in the first movement of J. C. Bach’s Symphony in E-flat major op. 9 no. 2, or Haydn’s Quartet in G major, op. 77 no. 1.⁵⁹

In each cited case, the apparent sonata-principle effect results from the movement’s engagement with a different, more complex type of compositional challenge. Here we need to separate out two axioms potentially suggested by Rosen’s argumentation. The first is that masterworks will not provide a “redundant” tonic statement of an S- or C-module in the development only to sound it again in the recapitulation. The second, the converse of the first (and closer to Rosen’s literal claims), insists that musical works of quality will not omit recapitulatory tonic statements of previously heard S- or C-modules unless there is a compensating tonic statement of the missing module elsewhere. The first seems generally correct; the second, more doubtful.

The first proposes that individual secondary-theme- or closing-modules (presented out-of-tonic in the exposition) normatively appear only once in the tonic key later in the piece—disregarding, of course, any double appearance effected by a repeat *en bloc* of the entire development and recapitulation. This single tonic appearance is usually unproblematic, occurring in the designated slot within the recapitulation. On rare occasions, however, an individual module or two might be displaced to a premature position—within the development. (Were both a developmental and a recapitulatory tonic statement of S or C permissible, then the presence of a developmental tonic S-module in

58. Rosen, *The Classical Style*, rev. ed., 73, 74.

59. Rosen, *Sonata Forms*, rev. ed., 288.

Op. 64 no. 3 would account for nothing. Rosen, we remember, implied that at least one of its functions there was to obviate the need for a tonic presentation of it in the recapitulation.) While this proposition may need more rigorous testing, it seems to be by and large valid. Any exception to it should be regarded as deformational—a purposeful breaking or overriding of the norm.⁶⁰

Although such considerations might at first strike one as esoteric, relevant only to the rarest compositions, grasping what lies behind them repays the effort: one is brought closer to some of the mainsprings of the sonata as a genre. Notice, for instance, that this “once-only” norm pertains principally to post-medial-caesura material from the exposition (its second half, S and C), not to the primary-theme (P) or transition (TR) zones. It is well known that a fleeting return to the tonic (often following a brief P-based passage beginning in V, although other options were also available) is common toward the opening of a sonata’s development, as Heinrich Christoph Koch famously noted.⁶¹ In practice, notwithstanding Koch’s directing our attention to other possibilities, this early tonic-station is often P-based and part of an initial sequential sweep of key-changes—typically striking us today as more of an apparent tonic than a real one.⁶² A more provocative tonic-key statement of the beginning of the primary theme, P¹—often claimed (though not without controversy) as a false or deceptive recapitulation—can also occur later in the development, only to be aborted by reverting back into developmental keys and textures.⁶³ While

60. Once again, in proposing this as a strong developmental guideline, one should distinguish between fleeting allusions on only apparent tonics (ones that appear, for instance, as one of several harmonic-notch levels in a broad sequence or circle-of-fifths motion [cf. nn. 55 and 61]—or fleetingly emergent tonics that are secondary to some larger harmonic unfolding) and ones that are secured with perfect authentic cadences in the tonic.

61. This is the more normal situation of the “apparent tonic” mentioned by Adrian: see note 55 above. Koch’s remarks about the possible early appearance of the tonic in the development (his “first period” of the “second part” of the form) are to be found in vol. 3, sec. 102 of his *Versuch einer Anleitung zur Composition* (1793; reprint, Hildesheim: Georg Olms, 1969), 307–8. Koch did not insist that such a tonic return following a sounding of P or a P-variant in V be accomplished with the P-idea itself, however: indeed, according to Koch it (normally? or at least in “die erste und gewöhnlichste Bauart dieses ersten Perioden”) occurs “vermittelst eines andern melodischen Theils.” Cf. the translation in Heinrich Christoph Koch, *Introductory Essay on Composition: The Mechanical Rules of Melody, Sections 3 and 4*, trans. Nancy Kovaleff Baker (New Haven: Yale University Press, 1983), 200: After beginning a development in V, often based on the “main melodic idea” or a variant, “it either modulates back into the main key by means of another melodic idea, and from this to the minor key of the sixth, or also to the minor key of the second or third.”

62. These early tonic-moments, often sounded with the P-idea, were noted and given the unfortunate label of “premature reprises” or “incomplete” recapitulations in a much noted article of Oliver Strunk, “Haydn’s Divertimenti for Baryton, Viola, and Bass,” *The Musical Quarterly* 18 (1932): 216–51. As is well known, much discussion of the phenomenon—and the term—has ensued.

63. Peter A. Hoyt disputes the claim of the “false recapitulation” interpretation of such passages in “Haydn’s ‘False Recapitulations,’ Late Eighteenth-Century Theory, and Modern

such occurrences—which turn up notably in some of Haydn’s quartets and symphonies in the late 1760s and early 1770s—may sometimes seem to invite modestly compensatory adjustments in the recapitulation, in the majority of cases they do not forestall a subsequent (second) tonic-P¹ appearance to launch that recapitulation. Many instances have been tabulated and discussed by Mark Evan Bonds: they include the finale of Haydn’s Symphony No. 48 and the first movements of his Symphonies Nos. 41, 55, 91, and a few others.⁶⁴ Moreover, tonic keys in general—perhaps with P¹-related material, if not P¹ itself—can also appear in passing in the development as one of the links within a sequential chain without jeopardizing a second appearance of the tonic at the recapitulation’s onset (as, electrifyingly, in the first movement of the *Eroica*, measure 316).

Why does this tonic-redundancy option seem to be generically less available for S-material? Clearly, expositional materials and functions are neither

Paradigms of Sonata Form,” paper presented to the Department of Music, Yale University, 30 March 2001. Hoyt argues that even later developmental appearances of P¹ in the tonic were likely to strike eighteenth-century listeners as largely normative—and thus relatable to the convention of early tonic appearances of P¹, described by several eighteenth- and early nineteenth-century theorists as a stabilizing element. I am grateful to Professor Hoyt for providing me with a copy of this paper.

64. See, for example, the convenient listing in Mark Evan Bonds, “Haydn’s False Recapitulations and the Perception of Sonata Form in the Eighteenth Century” (Ph.D. diss., Harvard University, 1988), 346. The list is complicated by issues of judgment concerning the status of a “premature” reprise (Bonds’s “precursory recapitulation”) versus that of a genuine false recapitulation: as Bonds notes, deciding which is which is not always easy. (Cf. the differing view of Hoyt, n. 63 above.) Moreover, some of Bonds’s cited changes in some of the subsequent recapitulations (the claim is that they “excise from the true recapitulation some or all of the material already presented [in the] false recapitulation”) are perhaps subject to questioning and somewhat differing interpretations. (As always in Haydn, every case seems unique.) Nevertheless, the main point for us is this: P¹ or strongly related P-material had been sounded emphatically in the tonic in the development—and the same or strikingly similar P¹-material is also heard in the tonic again to initiate the recapitulation.

The bibliography on the issue of false recapitulations is vast, the problems, differing definitions, and controversies tangled and intense. Such discussions invariably center around appearances of P-material in the development, not appearances of S. Cf. Leonard G. Ratner (*Classic Music: Expression, Form, and Style* [New York: Schirmer, 1980], 229), who cites the first movement of Haydn’s “Lark” Quartet in D, Op. 64 no. 5, as a “well-known” example of a (tonic-key) “false recapitulation” (m. 105; true recapitulation at m. 142). Bonds, on the other hand, questioned its status as a false recapitulation, because “it never [subsequently] strays from the home key long enough to create a sense of false recapitulation” (“Haydn’s False Recapitulations,” 232). However one nuances the matter, it is clear, once again, that P¹ has appeared in the tonic in the developmental space and that it recurs again to launch the recapitulation—our main concern here. For other discussions of similar issues, see Rosen, *Sonata Forms*, rev. ed., 276–81 (using Haydn’s Symphony No. 55 as the touchstone); Elaine Sisman, *Mozart: The “Jupiter” Symphony* (Cambridge: Cambridge University Press, 1993), 50–52; Peter A. Hoyt, review of *Wordless Rhetoric: Musical Form and the Metaphor of the Oration*, by Mark Evan Bonds, *Journal of Music Theory* 38 (1994): 123–43, esp. 135; and idem, “Haydn’s ‘False Recapitulations.’”

interchangeable nor equally sensitive with regard to tonal concerns. It matters greatly which expositional themes are selected for developmental treatment, how they are ordered, and what the nature of their handling is. Above all, they are not connotationally neutral. Even in the development, thematic references should be heard as recalling what their past function was and what their future role is “destined” to be. The expositional function of the secondary theme is to carry out a generically obligatory process that produces the first satisfactory perfect authentic cadence in the second key: the nontonic point of essential expositional closure (EEC, which may be followed by a postcadential closing zone, C). This is a vital task allotted within the genre only to S-materials, whose deployment proposes how cadential matters “could” work out smoothly in the (transposed) recapitulation-to-come. For this reason, the S-modules are the most generically critical materials of the exposition. It is the parallel task of the recapitulation’s secondary-theme zone to begin—finally—on the tonic (normally, of course, the initial S-module, S^1 —or, more strictly labeled, $S^{1.1}$, suggesting more strongly the first element of a multimodular secondary-theme zone—will launch this process)⁶⁵ and, guided by the exposition’s precedent, to lead to the whole structure’s most important formal moment, the attaining of essential structural closure (ESC), the corresponding perfect authentic cadence in the tonic key.⁶⁶

Accordingly, as a hypothesis one might propose that to state $S^{1.1}$ in the tonic *anywhere* in post-expositional space is to signal that the S-drive toward the ESC has been triggered and is under way. Apparently to abort that tonic drive—if, for example, the tonic $S^{1.1}$ leads only to more development, as in Op. 64 no. 3—is to render the prematurely sounded tonic-modules of S unavailable for the recapitulation-to-follow. In this case the opening tonic S-modules are spent or used up in the development.⁶⁷ Again, we should

65. With regard to numeric labeling: In multimodular thematic zones—such as the S-zones—it can be helpful for later referential purposes, as here, to distinguish among the differing modules. The initial exponential integer 1 means, in effect, “on the way to the first perfect authentic cadence.” (Although S-space normally consists of a single cadential span to a satisfactory perfect authentic cadence, it can happen via one of several “EEC-deferral strategies” that that S-space may extend beyond the first perfect authentic cadence [PAC] in its key; hence S^2 labels are meaningful.) The integer after the decimal point identifies the position of a modular idea within that cadential span for either distinctively musical or subsequently referential purposes. This convention, along with the problematics of EEC-deferral, is discussed in Hepokoski and Darcy, *Elements of Sonata Theory*.

66. Once again, for a more expanded treatment of the EEC and ESC, the central generic goals of all sonatas, see Hepokoski and Darcy, *Elements of Sonata Theory*. (Cf. n. 53 above.)

67. One or two prominent exceptions may be noted from a later repertory—and with composers who have become celebrated for their generically transgressive decisions with regard to tonal matters within sonatas. In the first movement of Schubert’s Piano Sonata in B \flat , D. 960, a variant of a prominent late-expositional theme first heard in measure 80 in V—perhaps a closing theme (although under some analyses it could be construed as part of S)—recurs in the development in B \flat minor (m. 146) as one link of a chain of ongoing sequences. Notwithstanding the change of key signature back to two flats here, this music is not set up as a new launch, but only as

presume as a necessary heuristic that such an occurrence is a component of a composer's calculated plan, articulating a larger compositional, expressive, or programmatic purpose. There must be a reason for spending, and thereby extinguishing, those components of S. In the first movement of Haydn's Op. 64 no. 3, the developmental presentation of those initial tonic S-modules, canceling their later availability, served the more compelling purpose of permitting a clever emendation of a problematic exposition (the TMB situation) in the subsequent recapitulation.

To the above may be added four additional hypotheses:

1. The tonic S-modules sounded prematurely in the development will most likely be taken from the beginning of the secondary-thematic material: $S^{1.1}$, $S^{1.2}$, and so on. (This seems confirmed by the cases cited by Rosen.) The later segments of S, arbitrarily identifiable as $S^{1.3}$, $S^{1.4}$, and the like, will be kept apart from tonic statements in the development, although they may occur in other keys. Instead, tonic appearances of the later secondary-theme modules are reserved for the recapitulation proper, since they (and especially the final S-module) were the ones that had driven to and accomplished the requisite completion of the perfect authentic cadence in the exposition, the EEC. In the recapitulation these later, successful modules were the ones "promised" in the exposition to provide the complementary, tonic-key ESC in the recapitulation. In general, therefore, later S-modules are given a more privileged functional importance within the entire structure than are earlier ones.
2. Tonic S-modules in the development (tonic music about to be abandoned for further nontonic adventures) may carry indications of their own provisionality: they may be presented as imperfect or flawed, even to the point of being considered unsatisfactory or inadequate presentations of S^1 . When compared with their expositional versions, for instance, they might be sounded in the parallel mode (especially in minor) as a sign of fragility or instability. Similarly, the tonic S^1 -module might be sounded on an only apparent tonic or insufficiently prepared by a complementary dominant-lock and medial-caesura effect—an important manifestation of the tenuous status of this module. All of these situations occur in the first movement of Haydn's Op. 64 no. 3. Such strategies of clouded presentation contribute to the compositional

another way station (an apparent tonic) in the modulatory plan. More problematic, perhaps, is the bizarre I:PAC (!) using late-exposition material (usually cited as S, though it may actually be, in effect, C) in the opening movement of Berlioz's *Symphonie fantastique*, measure 200. Almost surely related to the opium-enhanced *fantastique* nature of the program, such a "forbidden" occurrence cannot be cited as an example of anything—except, of course, of Berlioz's complex and quirkily deformational treatment of norms. In both cases the music in question does return in the recapitulatory rotation.

psychology: while the “normal” tonic $S^{1.1}$ -module (one sounding much like the parallel passage in the exposition) should initiate the S-zone pathway to the ESC, this one, weakened in initial articulation or flawed from the outset, proves incapable of sustaining itself and doing so.

3. Because of the importance of the mission generically assigned to them (producing the EEC and ESC), S-modules in all sonatas are more sensitive than are P-, TR-, and C-modules. No matter where and in which key they occur, secondary-theme modules remind us of their formal purpose: their vectored drive to the EEC/ESC moments. Within the development $S^{1.1}$ -modules may be sounded in any number of keys, but once they light upon the tonic or its parallel-modal substitute—and especially if they are led into with an attention-drawing dominant preparation and MC-effect—they touch a trigger-mechanism that can set off (in terms of expectations) the tonic drive toward the ESC. The heightened generic sensitivity of S-material may suggest why developments in the decades around 1800 more often center around P/TR- and/or C-modules. To be sure, S-based developments may certainly be encountered (Mozart, Piano Sonata in F, K. 280/iii; Beethoven, Symphony No. 5 in C Minor, Op. 67/iv, and many others), but they are not as common as ones built around other materials. In some cases nontonic S-statements within developments may follow P/TR-material as they participate in a broader strategy of thematic rotation (Beethoven, Symphony No. 2 in D, Op. 36/i; Piano Sonata in F Minor, Op. 57/i [“Appassionata”]). In other cases they may carry a charged urgency or expectation with regard to their task of closure down the road; in still others they may remind us of a problematic feature of S within the exposition, as if deliberating on the difficulty.
4. As opposed to the *cadential* S-modules, which serve to produce the obligatory perfect authentic cadences in the exposition and recapitulation (EEC and ESC), any later closing-zone modules (C) are *postcadential*. Within the exposition and recapitulation they exist more as optional extras (often gratifying and expansive ones) than as generically required ideas. They constitute what Koch called an *Anhang* (appendix) and Anton Reicha referred to as *idées accessoires*.⁶⁸ Consequently, they do not participate in the same sensitivities and imperatives as does S-space, nor

68. Koch, *Versuch einer Anleitung zur Composition* (1793), 3:305: “Oft ist zwar nach der Cadenz desselben noch ein erklärender Periode angehängt . . . ; daher können wir ihn für nichts anders, als bloß für einen Anhang des ersten Perioden erklären” (trans. Baker in *Introductory Essay on Composition*, 199: “Following the cadence a clarifying period is often appended. . . . Thus it is nothing else than an appendix to the first period”). Anton Reicha’s term *idées accessoires* (following the close of the nontonic “seconde idée mère” of the “exposition des idées”) is found in his celebrated diagram of the *grande coupe binaire*, from *Traité de haute composition musicale* (Paris, 1826), 2:300. The diagram has been widely reproduced, for example, in Ian Bent, *Analysis* (New York: Norton, 1987), 20.

do they convey the same kinds of functional connotations within the development. As an “after-the-fact” surplus, their main generic role is to signal the elegant final ideas of a thematic rotation. Perhaps for this reason the isolated onset of tonic C-modules within the developmental space would be a rare occurrence indeed: I am presently unaware of any examples. Within the recapitulation, of course, all of the C-modules, when they exist at all, are nearly always reiterated intact in the tonic. In this respect they almost always follow Rosen’s “rule”—his version of the sonata principle. But if one or more of them happens to be omitted (as in Mozart’s Symphony No. 34 in C, K. 338/i, mentioned at the beginning of this essay), we might presume that the sonata principle being violated is of a slightly lower order than any such omission found in S-space. Such an omission within the closing zone is doubtless expressively important for other reasons that need to be individually examined. Nevertheless, even while the sonata principle holds in most cases, the compelling need for tonic resolution probably applies more strongly to “obligatory” S-space—especially to its final, cadence-producing modules—than to the subsequent, generically “optional” C-space.

The above hypotheses have not yet broached the central issue. What are the larger implications into which a composer can tap by sounding a tonic statement of S^{1.1} in the development—particularly in its early or middle portions? Here we must recall that sonatas with developments and full recapitulations (what Warren Darcy and I have called “Type 3 sonatas”) constituted only the most common formatting option of the period. Also available, among others, was the lesser-used but still important “binary” sonata (our “Type 2”). In this sonata-type the tonal resolution (“recapitulation”), following a central development, began at or around the S-theme in the tonic.⁶⁹ Lacking a full recapitulation, Type 2 (“binary”) sonatas are anything but unfamiliar to students of music of this period: yet although they have been much mentioned, they have been often misconstrued. And indeed, their general availability presented a wide array of options and overrides to the late eighteenth-century composer.⁷⁰ They are more commonly encountered in works by the Mannheim composers, by J. C. Bach, and by Mozart, than in those by Haydn. (The many Mozartean instances include Symphony “No. 1” in E♭, K. 16/i; Symphony “No. 6” in F, K. 22/i; String Quartet in D, K. 155/ii; Overture to *Il re pastore*, K. 208; Violin Sonata in D, K. 306/i; Piano

69. We have laid out the five “numbered” types of sonatas and treated each at length in *Elements of Sonata Theory*. In brief: the Type 1 sonata is what has been called by others the “sonatina” or “sonata without development”; Type 2 refers to “binary” sonatas without full recapitulations; Type 3 is the standard “textbook” sonata form; Type 4 encompasses sonata-rondo hybrids; and Type 5 comprises adaptations of sonata form within concertos.

70. A discussion of these, along with a bibliography and evaluation of past treatments of the subject, is available in *Elements of Sonata Theory*.

Sonata in D, K. 311/i; *Eine kleine Nachtmusik*, K. 525/iv; and dozens of others.)⁷¹ Since it was a standard generic option, however, Haydn could not have been unfamiliar with it—and a clear use of the Type 2 format is found in the finale of his “Trauer” Symphony, No. 44 in E minor.⁷²

An adequate discussion of the Type 2 sonata and the complexities engendered by its many variants is impractical here. For our purposes, we need primarily observe that to sound a *sufficiently prepared*, early- or mid-“developmental” appearance in the tonic of $S^{1.1}$, especially within a manifestly rotational development, was to suggest that the movement was renouncing the more ordinary Type 3 format—with full recapitulation—by declaring itself generically on behalf of the more compact Type 2 sonata. Encountering a strong launch of the secondary theme in the tonic midway through an otherwise typical development could trigger the expectation that tonic S-space would continue, that the Type-2 S-drive toward the EEC and any subsequent C-modules was now under way.⁷³ Any case in which the tonic S^1 (or $S^{1.1}$, $S^{1.2}$, etc.) was soon undermined to plunge back into nontonic development—such as occurs in the first movement of Haydn’s Op. 64 no. 3—must have suggested that the “sudden” proposal of a Type 2 (“binary”) sonata was being just as quickly revoked and that the now-evanescent Type 2 was being reconverted into the far more normative Type 3 (“textbook”) sonata. In this case, whatever tonic S-modules were sounded during the development would have to be regarded as a “false tonal resolution,” the much rarer (Type-2-grounded) analogue of the more common false recapitulation of Type 3.

Apart from the production of witty, local structural dislocations, the most far-reaching result of the false-tonal-resolution strategy was in most cases to render the relevant tonic S-modules no longer available for deployment in the

71. This is not the place to enter into the discussion of the misunderstanding of the so-called reversed recapitulation or mirror form—a claim (often incorrectly applied, for example, to K. 306/i and K. 311/i, whose codas begin with references to the P-theme in the tonic) that is historically unjustifiable. Again, the correction of this misunderstanding is presented in the section of *Elements of Sonata Theory* dealing with the Type 2 sonata. It might also be mentioned that Mozart’s structural label for the finale of K. 525, “rondo,” is problematic, at least from the point of view of post-nineteenth-century customary classifications of form. Notwithstanding the multiple recurrences of the main theme, the work, complete with repeat signs, is unmistakably a Type 2 sonata. (Cf. the Rondo in D for Piano, K. 485, a Type 3 sonata.)

72. Another example occurs in “Haydn’s” Sonata in A, Hob. XVI:5/i, although the dating and authenticity of the sonata have not been fully established. See A. Peter Brown, *Joseph Haydn’s Keyboard Music: Sources and Style* (Bloomington: Indiana University Press, 1986), 63–75; and László Somfai, *The Keyboard Sonatas of Joseph Haydn: Instruments and Performance Practice, Genres and Styles*, trans. Somfai and Charlotte Greenspan (Chicago: University of Chicago Press, 1995), esp. the summary listing on p. 354.

73. Also relevant here—and adding to the local clarity and/or nuances of the event—are the decisiveness of the MC preparation of the tonic S-launch, the mode of the tonic S-modules, and the extent to which any S-modules had already been sounded in the development. The strongest cases feature clear MC preparation, matching mode with the exposition’s S, and no prior sounding of S in the development.

Type 3 recapitulation: their “once-only” ticket had already been spent in the development. Presumably the larger purpose of such a sonata-type conversion was to provide generic support for a preplanned recapitulation that, for any number of reasons, was to proceed in a manner not strictly parallel with the exposition.⁷⁴ Rosen’s proposal that incomplete S-themes within recapitulations may be justified because the missing portion of the theme “has already appeared *at the tonic* in the development section” touches on the crucial point, but the reasoning proceeds in only one of two necessary directions. In addition to arguing by means of retrojection or backward glances (from an incomplete recapitulatory S to an earlier, developmental tonic appearance of the lacking material), we should also examine the matter from the opposite direction—from the presence of tonic S-modules in the development to the recrafted recapitulation. Generally similar concerns may also be brought to bear on Rosen’s two parallel cases from Haydn’s quartets, the first movements of Op. 50 no. 6 and Op. 77 no. 1, neither of which, for reasons of space, will be dealt with here. (As expected, each movement features engaging local quirks and implications that differ from those in the first movement of Op. 64 no. 3. Among others, these include secondary themes derived from primary themes, a situation that can involve issues of recapitulatory thematic compression, momentary aspects of uncertainty regarding which theme is being alluded to, and the like. Ultimately, however, a consideration of these movements casts us back to hypotheses and conclusions relatable to those outlined above.)

Having considered the first axiom (the normative prohibition of redundant tonic S-modules), we may now turn to the second, which proposes that incomplete S-spaces (and possibly C-spaces) within recapitulations will have produced the missing modules, in the tonic, in the development. At least when taken as an absolute claim, this axiom is more difficult to defend. Here we must be as clear as possible. According to the directionally “forward”

74. Another example of this occurs in Weber’s D-major Overture to *Oberon*. The exposition’s S-zone is subdivided (deformationally) into *two* complementary themes (“his and hers,” as we learn from the opera that follows), each with its own perfect authentic cadence in the dominant: S¹, begun *dolce* in the clarinet, at measure 65; S² in the strings at measure 81, leading to the EEC at measure 101. Toward the end of the development, the brief opening module of “his” theme, S¹, is treated to two statements, one *forte* on III (F[♯] major, m. 154), the other, fleetingly, on a very weakly established tonic plane (D major, first sounded as \flat VI of F[♯], m. 158—at best a *Schein-tonik* effect—then moving at once to V of D, harmonically interrupted to begin the recapitulation with P¹ in the tonic, m. 165). S¹ is subsequently omitted entirely from the recapitulation. S² alone—“her” theme—remains and (in compensation?) is treated to the “Weber apotheosis” procedure of a climactic *fortissimo* restatement, measure 183. One may presume that the appearance of S¹ on the D-tonal plane in measure 158 serves in part as a means to legitimize the preplanned disappearance of S¹ from the recapitulation, the better to launch the anticipated S² apotheosis—and that the underlying strategy of all this surely had some programmatic basis (“her” apotheosis? the victory of love? love finally “found”? etc.). Since the S¹ in question appears only as the second statement of the thematic incipit (that is, since it is insufficiently prepared), and since it occurs only at the end of the development, this particular case probably does not raise the “Type-2” expectation mentioned in the text.

axiom 1, in cases where $S^{1.1}$, $S^{1.2}$, and so forth are sounded in the tonic in the development of a Type 3 sonata (as a false tonal resolution), those modules will typically be lacking in the recapitulation. But more frequent exceptions are found to the directionally “backward” axiom 2, its converse: it is not always the case that incomplete S- and C-spaces in the recapitulation will have been prepared by compensatory tonic S- and C-modules in the development. Still, there are no ironclad laws within genres, only norms of differing strengths. Rare exceptions to either axiom might be encountered as sonata deformations. What is proposed here is that exceptions to the second appear to be significantly more recurrent than exceptions to the first. In either case, however, we are dealing with sonata-deformational practice—the overriding of strong generic norms.

It is axiom 2, of course, that Rosen was most insistent upon. To requote *The Classical Style*: “These [three quartet movements] are the rare cases in the Haydn quartets of material exposed in the dominant and missing from the recapitulation, and at each point we have seen that some form of tonic recapitulation has been provided.”⁷⁵ This is misleading: other Haydn quartets provide counterexamples that Rosen passed over in silence. These include the first movements of Op. 20 no. 3 in G minor and Op. 33 no. 3 in C major, both of which feature early S-modules in the exposition that do not recur in the recapitulation and that make no tonic appearance in the development—or anywhere else. It will be useful to touch briefly on these pieces here.

In the exposition of the (G-minor) first movement of Op. 20 no. 3, following an unusually treated medial caesura (mm. 24–26), the B \flat -major secondary-theme zone (mm. 27–ca. 94) is uncommonly discursive, changing its mind and direction at nearly every turn. This secondary-theme zone, S, is subdivisible into at least eight different modules (!), which one may distinguish with the labels $S^{1.1-1.8}$.⁷⁶ In the abbreviated recapitulation, Haydn omitted the first

75. Rosen, *The Classical Style*, rev. ed., 74.

76. $S^{1.1}$ (mm. 27–33; this is a “P-based/P-derived” idea, but it is nonetheless a somewhat contrasting theme to P); $S^{1.2}$ (mm. 34–44; ascent of the first violin into the stratosphere and subsequent redescend, followed by a nervously scratching, *forte* trill-figure in the viola in the last four bars); $S^{1.3}$ (mm. 45–50; repeated-figure, rat-a-tat *forte* imitations between paired strings); $S^{1.4}$ (mm. 50–60; a first-violin descent and decrescendo back to *piano*, followed by a move toward what seems to be a “surefire” cadence-formula that, surprisingly, misfires onto a deceptive cadence at the end); $S^{1.5}$ (mm. 61–64; the similarly misfired, *forte* reaction to $S^{1.4}$, now pushing toward a wrong-key PAC on F major [V/III!]) followed by a rest in all voices); $S^{1.6}$ (mm. 65–70; a *fortissimo* fanfare-call back to order, followed by questioning, *piano* reflections on $S^{1.5}$ in the solo violin); $S^{1.7}$ (mm. 71–77: an even more discursive excursion for the first violin); $S^{1.8}$ (mm. 78–94; final attempts to cadence, once more undermined with a deceptive cadence [m. 85] and recall of the earlier caesura-fill figure from mm. 25–26 [mm. 86–87], followed by a final drive into a diminished-seventh ditch [mm. 93–94]). There is no concluding perfect authentic cadence to this exposition: the downbeat of measure 87 hardly counts as a moment of closure. Thus there is no EEC and no closing theme.

One may also observe that Haydn presents the later modules in such a way as to suggest that they are experiencing difficulty in producing what Caplin has termed a normative “expanded

three modules (originally sounded in mm. 27–49)—a situation roughly similar to that which would be revisited in Op. 64 no. 3 (although in the present quartet's exposition we are not presented with the issue of apparent double medial caesuras). What we do find are recomposed variants of $S^{1.4-1.8}$ (mm. 199–251, originally sounded in mm. 50–94).

How could one account for the recapitulatory suppression of the $S^{1.1-1.3}$ -modules under Rosen's version of the sonata principle? One might argue that since in Op. 20 no. 3 we are provided with the last five modules (1.4–1.8) in the recapitulation, the S-section is presented “fairly completely” (to adopt Rosen's flexible wording)—or at least completely enough to make the point of large-scale reconciliation. This is attractive reasoning, in line with what I suggested might be understood as a synecdochic strategy. The point, though, is that in this case that strategy would be operative without regard to prior tonic S-appearances in the development—thus undermining Rosen's “dissonant-section” contention. Following another track, one might reason that since $S^{1.1}$ in Op. 20 no. 3 (mm. 27–33) shares an initial rhythm with the primary theme (the P-based S was a characteristic, of course, of Haydn), it was unnecessary or redundant to restate it “in place” in the recapitulation. But Haydn by no means always feels compelled to suppress his P-based S-themes in the recapitulation, and it is difficult to extract any generally binding principle regarding the recapitulatory implications of a P-based S. Moreover, although this $S^{1.1}$ is P-based, as a whole idea it is clearly distinguishable from P: it shares only an initial rhythm with P but not its literal contour after the third pitch. Even if we could clear this hurdle—although it is one that we should not clear—the fact remains that the substantially different $S^{1.2}$ - and $S^{1.3}$ -modules (mm. 34–44 and 45–50, neither based on P-material) do not appear in the recapitulation at all.

Alternatively, if the issues of reconciliation and tonal stability are to hold according to what Rosen sought to demonstrate about Op. 64 no. 3, we should expect to find $S^{1.1}$, $S^{1.2}$, and $S^{1.3}$ presented in the tonic, G minor, in the development or coda. But this does not happen. The coda is entirely P-based (mm. 252–70). And the development of the first movement of Op. 20 no. 3 is essentially rotational, with the expositional materials treated in the order: P (mm. 96–109); the P-based TR (mm. 110–20); the concluding measures only of $S^{1.2}$ (mm. 121–24 now articulating mostly V of Eb, as opposed to the original mm. 41–44, mostly on V/III); $S^{1.3}$ (mm. 125–29; I–V alternations in Eb; cf. mm. 45–50, I–V alternations in Bb); $S^{1.8}$ (mm. 130–40, still implying

cadential progression,” one in which cadential function is enlarged over “all four harmonic functions (initial tonic [often I^6 as a ‘conventionalized sign’], pre-dominant, dominant, and final tonic),” often associated, for example, with the familiar $\hat{3}-\hat{4}-\hat{5}-\hat{1}$ motion in the bass. On the general concept, see, for example, Caplin, *Classical Form*, 109–11; and idem, “The ‘Expanded Cadential Progression’: A Category for the Analysis of Classical Form,” *Journal of Musicological Research* 7 (1987): 215–57.

E♭ major).⁷⁷ In other words, the modules S^{1.2} (last portion only) and S^{1.3}, which will be “missing” from the recapitulation, appear in the development only within the orbit of E♭ major, not the G-minor tonic. S^{1.1} (albeit P-based) does not appear at all after the exposition, nor does the first half of S^{1.2}.

At this point one might seek to bail out of the difficulty by appealing to the apparently Conian principle of “closer relation.” In other words, one might claim that S^{1.2} and S^{1.3}, sounded in E♭ in the development—a fifth lower than their mediant-major statements in the exposition—somehow fulfill the sonata principle through their fifth-relationship to their original statement. But upon further reflection this “argument from fifth-transposition” can never be sustained, here or elsewhere.⁷⁸ If off-tonic developmental statements of expositional themes were capable of this substitute- (or second-level-) resolution function, the whole tonal rationale of normative recapitulatory logic would be jeopardized. The rickety “fifth-transposition” argument is best rejected as an inadequately considered patch for a defectively posited version of the sonata principle.⁷⁹

My own view is that the first movement of Op. 20 no. 3 is an extravagantly capricious piece, a paragon of purposeful disorder and distraction.⁸⁰ Time and again it juxtaposes the discursively prolix with the nervous compression of

77. What follows—seemingly outside of the rotational principle—is a backing up to the P-based TR-material (E♭, m. 141; cf. m. 110), which is almost immediately aborted and diverted into a new retransition to start the recapitulation proper (m. 165).

78. Cf. Rosen’s similar argument concerning a theme from the first movement of Mozart’s Sonata in C Minor, K. 457, in *Sonata Forms*, rev. ed., 288. But the theme in question (stated “in the development section at the subdominant”) occupies an expositional space (m. 23, as part of a thematized TR) that is prior to the more convincingly “real” MC of the movement, which occurs in measure 34. As suggested in the preceding section of this essay, pre-MC material is not under the same imperatives of tonal resolution as post-MC material. (Cf., again, Beethoven, Op. 10 no. 3/i, mentioned in nn. 8, 10, and 38 above.) More precisely, K. 457/i features the lesser-used type of TMB in which the first “apparent” MC is, in effect, no satisfactory MC at all, even though it leads to an “S-like” thematized TR, which indeed sets out in III. In short, from this perspective Rosen misidentified the Mozartean theme as part of “the second group.”

79. A more elaborate discussion of the “fifth-transposition” argument is provided in Hepokoski, “Back and Forth from *Egmont*” (n. 10 above; cf. n. 83 below).

80. A related—though not identical—issue was pursued by Elaine Sisman in her paper “C. P. E. Bach, Beethoven, and the Labyrinth of Melancholy,” delivered at the Sixth-sixth Annual Meeting of the American Musicological Society, Toronto, 2 November 2000; an expanded version was presented to the Department of Music, Yale University, 29 November 2000. (I am grateful to Professor Sisman for providing me with a copy of this paper.) Sisman related a number of minor-mode works or sections thereof—normally in slow tempo—to the contemporary discourse surrounding melancholia. Frequently associated with this sadness of temperament were such features as a studious frame of mind, extreme mental acuity and memory, a high degree of self-absorption (though occasionally leading to apparent surface disorder), and occasionally a labyrinthine convolution of thought process. Many eighteenth-century writers took pains to distinguish it from the extreme of genuine madness (typically understood as more raving or violent), but from time to time, as Sisman mentions, room was permitted for melancholy to slide into such states as “melancholy madness.”

“wishing to get on with it, to accomplish cadential goals.” The exposition is staged as a connoisseur’s comedy, proceeding beyond the pale in terms of inefficiency and nonconcentration. The exposition’s eventual lack of closure (ending on $\text{vii}^{\circ 4}/\text{i}$ in m. 94) only heightens the eccentricity of the musical thought. The goal of this sonata form as a whole seems to be to curb the immoderations of the exposition’s S. Toward that end the development scrubs away some of the apparently superfluous material, encouraging the recapitulation to present a relatively more normatively arrayed or disciplined set of materials (although eccentricities still abound in it). In this reading, the rotational pattern and a larger governing idea of trimming back the expositional overgrowth predominate over the promptings of the sonata principle per se. Tonal reconciliation of all expositional elements is not the central point.

Some of the same issues are also at hand in the more compact first movement of the Quartet in C, Op. 33 no. 3 (“Bird”).⁸¹ The exposition is typically concentrated in its treatment of motives. Following the articulation of the medial caesura (m. 26), S, reinvigorating P-motives, begins in measure 27. S proper unfolds as a sentence whose presentation modules constitute a varied basic idea ($S^{1.1}$, m. 28, G major; $S^{1.2}$, m. 30, with a decay into G minor, “lights out”), and whose apparent continuation, itself sentential (perhaps even something of a “presentational” rebeginning), may be subdivided into $S^{1.3}$ (m. 32, struggling in the snares of G minor) and the additionally continuational and ultimately cadential $S^{1.4}$ (m. 36, a G-major breaking-free), leading to the first satisfactory perfect authentic cadence, the EEC, in G major, at measure 42. (See Ex. 12a–b.) S is followed by P-based C-material (mm. 43–56, merging into a retransition). Missing from the recapitulation are $S^{1.1}$ and $S^{1.4}$, and neither of these thematic ideas is given a tonic presentation in the development. In fact, the thematic configuration of the major-mode version of $S^{1.1}$ (cf. m. 28) is never heard after the exposition, although that of the minor-mode variant $S^{1.2}$ (cf. m. 30) does appear several times.

81. As they are in yet another example, Mozart’s C-major Overture to *Così fan tutte*, K. 588, which omits the S-block entirely from the recapitulation, while the development is constructed from whirling *nontonic* cycles of TR- and S-material. The omission of S from the recapitulation, therefore, is unanticipated by any earlier *tonic* appearance of S, although that theme is certainly swirled about in other keys in the development. In brief: most unusually, the overture comprises the witty, frenzied switches of position and rearrangements of a limited number of breathless thematic modules. (Here the potential for analogies to the plot-to-follow is obvious.) In the exposition, following an MC at measure 57, S emerges in G minor (v) at measure 59—pasted together from materials sounded in preceding modules; the G-major EEC (and hence end of S-space) most probably should be taken to occur at measure 79. The development, beginning with the cadence in measure 95, is first treated to three subrotations of TR- and S-modules: (1) measure 95, G major (V/C) to A minor-major (hint of V/d); (2) measure 115, A major (V/d) to G minor-major (V/c); (3) measure 133, G major (V/C) to C major (for TR only, also construable as V/f) to a descending-circle-of-fifths treatment of S (D_5° , g, C_5° , F, etc.). The music emerges out of the subrotational cycles and onto F major at measure 149. Any further discussion of this extraordinarily complex overture—including the structural role of its “*Così fan tutte*” cadence, first sounded in the introduction—would demand more space than is practical here.

Example 12 Haydn, String Quartet in C Major, Op. 33 no. 3, first movement: (a) mm. 1–6; (b) mm. 22–42

(a)

Allegro moderato

Violino I

Violino II

Viola

Violoncello

4

f

f

f

f

Example 12 continued

(b)

22

Musical score for measures 22-27. The score is in 4/4 time and consists of four staves: two treble clefs (top and second) and two bass clefs (third and bottom). The music features a complex texture with multiple melodic lines and rhythmic patterns. Measure 22 starts with a treble clef staff playing a series of eighth notes, while the bass clef staff plays a similar pattern. The second treble clef staff has a more active line with various note values and rests.

28

Musical score for measures 28-32. This section continues the complex texture from the previous system. The top treble clef staff features a prominent melodic line with many sixteenth and thirty-second notes. The bass clef staff provides a steady accompaniment with eighth notes. The second treble clef staff has a more active line with various note values and rests.

33

Musical score for measures 33-37. This section continues the complex texture. The top treble clef staff features a prominent melodic line with many sixteenth and thirty-second notes. The bass clef staff provides a steady accompaniment with eighth notes. The second treble clef staff has a more active line with various note values and rests. Dynamic markings *fz* are present in measures 35 and 37.

38

Musical score for measures 38-42. This section continues the complex texture. The top treble clef staff features a prominent melodic line with many sixteenth and thirty-second notes. The bass clef staff provides a steady accompaniment with eighth notes. The second treble clef staff has a more active line with various note values and rests. Dynamic markings *fz* are present in measures 38, 40, and 42. The word *semplce* is written above the second treble clef staff in measure 40. The letter *(p)* is written below the second treble clef staff in measures 40 and 41, and below the bottom bass clef staff in measure 41.

Intricate in its tightly knit logic, the movement presents additional complications. The major-minor tussles of S in the exposition, for instance, are pacified in the recapitulation's more consistently major-mode recasting of some of its material (as if the minor-mode challenges had been exorcised in the development). But the recapitulation's strategy of recomposition is even more substantial than that: lacking a clearly articulated medial caesura, the recapitulation—at least up to the ESC in measure 138—is shaped as a continuous block of *Fortspinnung*, not as a two-part reprise with the traditional gap between TR- and S-space. Haydn thus recrafted the recapitulation as a structural block without an S-theme proper, even though some of the S-modules from the exposition return here. This is not an unfamiliar procedure in Haydn's works: two-part expositions could be reconceived as continuous recapitulations and vice versa. Perceiving this most basic of alterations is the first step toward any analytically sufficient consideration of such a work. Encountering such reconceived treatment, one should not be surprised to discover that certain S-materials fail to make an appearance in the recapitulation.⁸²

Moreover, in the first movement of Op. 33 no. 3, the development is laid out as two small-scale rotations. The first places the following in an ordered succession: P (m. 60, F major), S^{1.2} motives (m. 66, a weak B♭ major, G minor, to V/A minor), C (m. 74, A minor), and the retransition (m. 85). The second: C-minor (tonic) P-references (m. 88, cadencing on V of A minor, reinforced with fermata-pauses), S^{1.2} motives again (m. 99, mostly E minor), and an essentially intact S^{1.4} (m. 103, E minor, with a strong perfect authentic cadence in that key in m. 108). At least in this C-major quartet—this is anything but an expectation within the style—the development's strong E-minor (nontonic) cadence on S^{1.4} material seems to have the effect of liquidating this

82. An early example in Haydn may be found in the first movement of Symphony No. 6 in C (“Le matin”)—a two-part exposition (with a wittily “blank” S¹-module) subsequently converted into a continuous recapitulation. The opposite situation may be found in the first movement of Symphony No. 75 in D (problematically mentioned by Rosen, as quoted above, in support of his sonata-principle “rule”). This highly subtle movement—with a thoroughly reconstrued and re-ordered recapitulation—does not lend itself to quick descriptions. The exposition is best regarded as continuous (lacking a fully effective medial caesura proceeding to a clear S): a blocked-caesura effect occurs in measure 49, and a thematized, expanded caesura-fill, measures 49–55, brings us, unexpectedly, to a perfect authentic cadence in V at measure 56. This structural cadence occurs late in the exposition, some 73 percent of the way through, and consequently what follows (mm. 56–68) is doubtless *forte* closing material, C. Thus the exposition problematizes the expected moments of MC, caesura-fill, and S: what we initially construe as a non-normative caesura-fill (pre-S) itself takes on the (S) task of producing the EEC in measure 56. The recapitulation compensates by providing a more proper MC (m. 134, a perfect authentic cadence in I) and turning the earlier caesura-fill into an exceptionally discursive S-space (mm. 135–59), with ESC at measure 159. (Additionally, the exposition's C¹-material is resituated within the recapitulation's TR-space.) For the conceptual background and related issues concerning the blocked-caesura effect in general, see Hepokoski and Darcy, “The Medial Caesura,” 123 n. 18; and idem, *Elements of Sonata Theory*.

thematic unit for the rest of the piece.⁸³ As a result, at the point in the recapitulation when $S^{1.4}$ —the all-important S-cadential module—is expected (following references to $S^{1.3}$, mm. 129–34), Haydn breaks off the train of thought, as if it were quizzical, uncertain, stranded in a blind alley, and substitutes a placidly banal new idea to descend to the perfect authentic cadence that marks the ESC (mm. 136–38; Ex. 13). We might suppose that committed sonata-principle adherents would seize on the “latent relations” between the new “banality” and the omitted $S^{1.4}$ (both feature a prominent $\hat{6}-\hat{1}$ descent common to several cadential points in this movement and a similar contour, for instance, and both involve differing amounts of *dolce* material), thereby arguing that $S^{1.4}$ is not really absent. In a sense, this is not incorrect. (It is also true that the “banal” figuration of measures 136–38 had cropped up fleetingly in earlier moments of the movement.)⁸⁴ But the new theme lacks both the expanse and the vivid, idiosyncratic shaping of the seven-measure $S^{1.4}$. More to the point, such an attempt at normalizing the unusual would serve only to desensitize us to what actually occurs here: most uncommonly, it is the end of the S-material, that zone’s most functionally crucial module, that finds itself empty-handed, in need of the witty quick fix to be supplied by all-purpose, stereotypically at-hand replacement material.

Here in Op. 33 no. 3, much (but not all) of the S-material recurs in the tonic within a continuous recapitulation. Once again, we have an application of the infrequent synecdochic strategy, whereby a salient part of S-material can appear in the recapitulation to stand for the whole of S, which has been shortened and reshaped for other, overriding compositional reasons. Moreover, we also find here an instance of what might be called the substitution strategy, in which one or more S-modules are replaced in the recapitulation by a functional counterpart. Again, it is the analyst’s duty not to blur the differences between the original module and its substitute but rather to

83. One potentially engaging observation—also relevant, perhaps, to the G-minor/E \flat -major tonal situation discussed earlier in Op. 20 no. 3—might be anticipated from current neo-Riemannian theorists, who might wish to point out that E minor and C major stand in an “L” (*Leittonwechsel*) relation to each other. (In other words, E minor could be transformed into C major through a simple half-step $\hat{5}-\hat{6}$ shift.) As such, to use Richard L. Cohn’s terminology, they are “adjacent harmonies” that are hexatonically and “smoothly” related. The main question, though, is to what extent in this period the one can be reasonably understood as substituting for the other. For the basic terminology and concepts, see, for example, Cohn, “Maximally Smooth Cycles, Hexatonic Systems, and the Analysis of Late-Romantic Progressions,” *Music Analysis* 15 (1996): 9–40; idem, “As Wonderful as Star Clusters: Instruments for Gazing at Tonality in Schubert,” *19th-Century Music* 22 (1999): 213–32; and the entire issue of *Journal of Music Theory* 42, no. 2 (1998), devoted to neo-Riemannian theory. For my brief reply within the Haydn-Mozart-Beethoven repertory, see “Back and Forth from *Egmont*” (n. 10 above).

84. See, for example, the viola in measure 19 and the cello in measure 25. Both of these instances, however, occur within the transition (TR), and such cross-zone correspondences between TR and the new substitute for $S^{1.4}$, while both notable and engaging, are not properly adduced within any sonata-principle argument.

Example 13 Haydn, String Quartet in C Major, Op. 33 no. 3, first movement, mm. 128–38

heighten them and to ask why the composer passed up the easier, strict-transpositional route in order to pursue an unexpected path.⁸⁵ However infrequently they might be invoked, the synecdochic and substitution strategies, linked with the larger motivating reasons that they were employed at all, can override the sonata principle. True, we might elect to regard the result as purposefully deformational—intentionally transgressive of the norm—but we

85. One of the most provocative (and familiar) instances of (very rare!) cadential-module suppression and substitution occurs in the first movement of Beethoven's Piano Concerto No. 4 in G, Op. 58. Here the module producing the point of essential expositional closure, the EEC, in what is effectively the exposition proper is never sounded again. (We may call it $S^{1.4}$, mm. 180–88.) This is material that, in this case, is situated in the second orchestral tutti, a compositional zone that by no means customarily has this EEC-producing function. More often, the closure associated with the EEC is produced in the solo-exposition proper. (Cf. the parallel role of $S^{1.4}$ in the initial tutti-rotation, mm. 60–68.) In the recapitulation (following the unusually placed cadenza), we find it displaced by another one of the S-modules (m. 347)—a wide-eyed, dreamily “time-stopping” lyrical theme that has been increasingly encroaching upon rotational space since it was first sounded in measure 50. Coming out of the cadenza—and counter-generically—it is this highly personalized new theme that is granted the central task of producing the ESC in measure 356. This entire process of the eventual replacement of $S^{1.4}$ is a central feature of the large-scale unfolding of this movement.

ought to resist explanations that impede us from noticing, and valuing, the exceptional nature of these instances.

Even while the more modest versions of the sonata principle—those restricting themselves to the expectations of recapitulatory restatements of expositional material (more properly, of post-MC expositional material)—are generally upheld within individual movements, they do not apply in all instances. We should not think that the “rule” is inviolable, much less try to make fractious exceptional cases conform to it. Instead of normalizing puzzling occurrences, we should seize with relish upon “what doesn’t fit” and assume that pursuing the strangeness of that moment will lead us more deeply into the work’s driving compositional logic. And we should be prepared within all works—not only in those of the constantly surprising Haydn—to think through genre-implicated compositional problems on a deeper level, not to smooth out the works into blandly flawless “unities” and “reconciliations” but to savor their spiky eccentricities, to delight in the problematics of their discourse.

Envoi

In questioning the sonata principle’s range of application within compositions from the decades around 1800, much of the preceding essay has dealt with hard cases requiring close analysis—most notably some provocative Haydn quartets. Such scrutiny was unavoidable: it is in just such problematic instances that variants of this principle have typically been invoked, and it is only when we pause to look more carefully at the music that we may perceive the principle’s shortcomings. With that in mind I have forgone discussions of more obvious transgressions of it, notably those cases in which the post-medial-caesura portion of the recapitulation (the secondary and closing themes) either veers away from the tonic altogether or is sounded entirely in the “wrong” key. These would include, among others, the bizarre first movement of Mozart’s String Quartet in D Minor, K. 173 (another example, with the first movement of Haydn’s Op. 20 no. 3, of an instrumental representation of extreme melancholy or even madness?),⁸⁶ the E-major slow movement of Beethoven’s Piano Trio in G Major, Op. 1 no. 2, and, most famously, the F-minor *Egmont* Overture, in whose recapitulation S and C are made to appear in VI, D \flat major (presumably as a programmatic image of tonal dislocation or narrative failure).

Each of these is an instance of the broader phenomenon that I have called the *nonresolving recapitulation*, a striking and subsequently influential sonata deformation to which I have devoted a separate essay.⁸⁷ It is self-evident that

86. See note 80 above.

87. Hepokoski, “Back and Forth from *Egmont*.”

such extreme treatments of recapitulatory space violate the presumed sonata principle, and, moreover, they bring up related issues, such as the function of corrective codas (often extended in length) outside of sonata space, the cardinal importance of the rotational concept in grasping such structures, and the possible overlay of programmatic, narrative, or other representational features onto the unfolding of sonata rhetoric. As I suggest in that separate essay—complementary to the present one—the nonresolving recapitulation would take on a significant expressive role in many pieces of the later nineteenth and early twentieth centuries. Entering into that discussion here—or, for that matter, entering into the tonal and thematic problems in Schubert’s recapitulations (whose seeming secondary themes are sometimes never sounded in the tonic at all)—would lead us away from our present, “classical”-centered discussion, where the adducing of the sonata principle is most likely to be encountered.

Similarly, I shall only remind us here of other possibilities for more spectacular violations of the principle: the omission of the secondary and closing themes altogether in the recapitulation of Mozart’s *Idomeneo* Overture, which seems literally to compose out the recapitulation’s destruction or collapse in a violent and much extended transitional- (pre-MC-) space (mm. 115–end); or the related slow-movement variants, in which the apparent recapitulation of a sonata-ABA’ hybrid brings back only pre-medial-caesura material (P or P and TR) in its recapitulatory space, moving directly into a coda (with perhaps only indirect S-allusions), as in the second movements of Haydn’s string quartets Op. 33 nos. 5 and 6 and Mozart’s string quartet K. 575.⁸⁸ If only for practicality’s sake we shall pass by such eccentric or deformational works in order to proceed to a wrapping up of the general issue.

As a guideline for analysis the sonata principle as it has come down to us in its various formulations was a mid- to late-twentieth-century answer to the problem of confronting the multiplicity of sonata practice in an analytical environment that (in part as a corrective to earlier simplicities) was encouraged to downplay the structural importance of themes and thematic arrangement in favor of tonal relations. One of its attractions was its power to combine quick answers to anomalies of thematic disposition with the simultaneous reinforcement of its own presuppositions. On the one hand, it provided more latitude for where and in which contexts, structural spaces, or keys a sonata’s thematic modules could appear or not appear, thus rendering any further wondering about the matter pointless. On the other hand, it could cite any exception as

88. See, for example, the discussion of such recapitulations in Hepokoski and Darcy, *Elements of Sonata Theory*. Caplin (*Classical Form*, 216) calls this the “truncated recapitulation” and cites other examples of it on p. 217, in table 14.1. For some potential ramifications in the nineteenth century, see Elaine R. Sisman, “Brahms’s Slow Movements: Reinventing the ‘Closed’ Forms,” in *Brahms Studies: Analytical and Historical Perspectives*, ed. George L. Bozarth (Oxford: Clarendon Press, 1990), 79–103.

demonstrating the more fundamental assumption, namely, that thematic treatment and disposition were at best secondary, perhaps even arbitrary, matters in this repertory.

As always, nuance is everything, and one should not overstate one's claims. It is true—as suggested at the outset—that several recent writers have relativized Cone's sonata principle, reinterpreted it, or considered it as only one factor among others in the construction of sonatas. One might further expect that, when pressed, most thoughtful commentators on “the Classical style” would always have backed away from any absolutist claim regarding this principle. For some readers the preceding discussion may seem to have elevated the sonata principle into something more rigid than it has generally been taken to be. Some may object that I have overconstructed it as a straw figure to knock down. That has not been my intention.

Instead, I have sought only to call attention to an internal discrepancy that surfaces regularly in our commonly accepted analytical methods. When the sonata principle is isolated as a topic in itself (as here)—called forth for examination—it stands revealed as reduced in its utility and reliability as an interpretive tool within the repertory. As many have doubtless realized, to interrogate the principle *qua* principle is to become rapidly aware of its limitations and the early, exaggerated claims made on its behalf. Nevertheless, in the heat of practical analysis—when the focus is on devising an efficient explanation of the anomalous individual piece—the sonata principle is often too enticing a tool to resist. One grabs quickly into the analytical toolbox, and there it is, a largely unexamined piece of intellectual equipment packed into the box by someone else. True, a twist of the sonata-principle wrench temporarily closes off the leakage into further inquiry. But whenever we find it used as that single, simple tool—as an offhand citation to provide the quick analytical *aperçu* or the broad, overall principle of justification—we should step back, be cautious. We have surely learned by now that in the most interesting cases it is inadequate to the task at hand.

Works Cited

- Adrian, Jack. “The Function of the Apparent Tonic at the Beginning of Development Sections.” *Intégral* 5 (1991): 1–53.
- . “The Ternary-Sonata Form.” *Journal of Music Theory* 34 (1991): 57–80.
- Barford, Philip T. “The Sonata-Principle: A Study of Musical Thought in the Eighteenth Century.” *The Music Review* 13 (1952): 255–63.
- Bent, Ian. *Analysis*. The Norton/Grove Handbooks in Music. New York: Norton, 1987.
- Bonds, Mark Evan. “Haydn's False Recapitulations and the Perception of Sonata Form in the Eighteenth Century.” Ph.D. diss., Harvard University, 1988.
- . “The Paradox of Musical Form.” In *Wordless Rhetoric: Musical Form and the Metaphor of the Oration*, by Mark Evan Bonds, 13–52. Cambridge: Harvard University Press, 1991.

- Brown, A. Peter. *Joseph Haydn's Keyboard Music: Sources and Style*. Bloomington: Indiana University Press, 1986.
- Bushler, David. "Harmonic Structure in Mozart's Sonata-Form Developments." In *Mozart-Jahrbuch 1984/85*, 15–24. Kassel: Bärenreiter, 1986.
- Caplin, William E. *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven*. New York: Oxford University Press, 1998.
- . "The 'Expanded Cadential Progression': A Category for the Analysis of Classical Form." *Journal of Musicological Research* 7 (1987): 215–57.
- Cohn, Richard L. "As Wonderful as Star Clusters: Instruments for Gazing at Tonality in Schubert." *19th-Century Music* 22 (1999): 213–32.
- . "Maximally Smooth Cycles, Hexatonic Systems, and the Analysis of Late-Romantic Progressions." *Music Analysis* 15 (1996): 9–40.
- Cole, Malcolm S. "Czerny's Illustrated Description of the Rondo or Finale." *The Music Review* 36 (1975): 5–16.
- . "Sonata-Rondo, the Formulation of a Theoretical Concept in the Eighteenth and Nineteenth Centuries." *The Musical Quarterly* 55 (1969): 180–92.
- Cone, Edward T. *Musical Form and Musical Performance*. New York: Norton, 1968.
- Einstein, Alfred. *Mozart: His Character, His Work*. Oxford: Oxford University Press, 1945.
- Haimo, Ethan. *Haydn's Symphonic Forms: Essays in Compositional Logic*. Oxford: Clarendon Press, 1995.
- Hepokoski, James. "Back and Forth from *Egmont*: Beethoven, Mozart, and the Nonresolving Recapitulation." *19th-Century Music* 25 (2002): 127–54.
- Hepokoski, James, and Warren Darcy. *Elements of Sonata Theory: Norms, Types, and Deformations in the Late-Eighteenth-Century Sonata*. New York: Oxford University Press, forthcoming.
- . "The Medial Caesura and Its Role in the Eighteenth-Century Sonata Exposition." *Music Theory Spectrum* 19 (1997): 115–54.
- Hinrichsen, Hans-Joachim. "Sonatenform, Sonatenhauptsatzform." 1996. In *Handwörterbuch der musikalischen Terminologie*, edited by Hans Heinrich Eggebrecht, 1–20. Stuttgart: Steiner, ca. 1972–.
- Hoyt, Peter A. "Haydn's 'False Recapitulations,' Late Eighteenth-Century Theory, and Modern Paradigms of Sonata Form." Unpublished paper presented to the Department of Music, Yale University, 30 March 2001.
- . Review of *Wordless Rhetoric: Musical Form and the Metaphor of the Oration*, by Mark Evan Bonds. *Journal of Music Theory* 38 (1994): 123–43.
- Keller, Hans. "The Chamber Music." In *The Mozart Companion*, edited by H. C. Robbins Landon and Donald Mitchell, 90–137. New York: Norton, 1956.
- Kerman, Joseph. "Tovey's Beethoven." In *Write All These Down: Essays on Music*, by Joseph Kerman, 155–72. Berkeley and Los Angeles: University of California Press, 1994. Originally published in *The American Scholar* 45 (1975–76): 795–805; also in *Beethoven Studies* 2, ed. Alan Tyson (London: Oxford University Press, 1977), 172–91.
- Kirkpatrick, Ralph. *Domenico Scarlatti*. Princeton: Princeton University Press, 1953.
- Koch, Heinrich Christoph. *Introductory Essay on Composition: The Mechanical Rules of Melody, Sections 3 and 4*. Translated by Nancy Kovaleff Baker. New Haven: Yale University Press, 1983.
- . *Versuch einer Anleitung zur Composition*. 3 vols. 1793. Reprint, Hildesheim: Georg Olms, 1969.

- LaRue, Jan. *Guidelines for Style Analysis*. New York: Norton, 1970. 2d ed., Warren, Mich.: Harmonie Park Press, 1992.
- Mellers, Wilfrid. *The Sonata Principle (from c. 1750)*. Fair Lawn, N.J.: Essential Books, 1957.
- Morgan, Robert P. "Coda as Culmination: The First Movement of the 'Eroica' Symphony." In *Music Theory and the Exploration of the Past*, edited by Christopher Hatch and David W. Bernstein, 357–76. Chicago: University of Chicago Press, 1993.
- Newman, William S. *The Sonata in the Classic Era*. New York: Norton, 1963. Rev. ed., 1972.
- Plantinga, Leon. *Beethoven's Concertos: History, Style, Performance*. New York: Norton, 1999.
- Ratner, Leonard G. *Classic Music: Expression, Form, and Style*. New York: Schirmer, 1980.
- . "Harmonic Aspects of Classic Form." *This Journal* 2 (1949): 159–68.
- Réti, Rudolph. *The Thematic Process in Music*. New York: Macmillan, 1951.
- Rosen, Charles. *The Classical Style: Haydn, Mozart, Beethoven*. New York: Norton, 1971–72. Rev. ed., 1997.
- . *Sonata Forms*. Rev. ed. New York: Norton, 1988.
- Schenker, Heinrich. *Free Composition*. Edited and translated by Ernst Oster. 2 vols. New York: Longman, 1979. Originally published as *Der Freie Satz* (Vienna: Universal Edition, 1935).
- Schoenberg, Arnold. *Fundamentals of Musical Composition*. Edited by Gerald Strang and Leonard Stein. London: Faber, 1967.
- Seiffert, Wolf-Dieter. *Mozarts frühe Streichquartette*. Munich: Fink, 1992.
- Sisman, Elaine R. "Brahms's Slow Movements: Reinventing the 'Closed' Forms." In *Brahms Studies: Analytical and Historical Perspectives*, edited by George L. Bozarth, 79–103. Oxford: Clarendon Press, 1990.
- . "C. P. E. Bach, Beethoven, and the Labyrinth of Melancholy." Paper presented at the Sixty-sixth Annual Meeting of the American Musicological Society, Toronto, 2 November 2000. Expanded version presented to the Department of Music, Yale University, 29 November 2000.
- . *Mozart: The "Jupiter" Symphony*. Cambridge: Cambridge University Press, 1993.
- Somfai, László. *The Keyboard Sonatas of Joseph Haydn: Instruments and Performance Practice, Genres and Styles*. Translated by László Somfai and Charlotte Greenspan. Chicago: University of Chicago Press, 1995.
- Strunk, Oliver. "Haydn's Divertimenti for Baryton, Viola, and Bass." *The Musical Quarterly* 18 (1932): 216–51.
- Tovey, Donald Francis. "Brahms's Chamber Music." 1929. In *The Main Stream of Music and Other Essays*, by Donald Francis Tovey, 220–70. Oxford: Oxford University Press, 1949.
- . *A Companion to Beethoven's Pianoforte Sonatas*. London: Royal Schools of Music, 1931.
- . "Haydn's Chamber Music." 1929. In *The Main Stream of Music and Other Essays*, by Donald Francis Tovey, 1–64. Oxford: Oxford University Press, 1949.
- . "Musical Form and Matter." 1934. In *The Main Stream of Music and Other Essays*, by Donald Francis Tovey, 160–82. Oxford: Oxford University Press, 1949.

- . “Some Aspects of Beethoven’s Art Forms.” 1927. In *The Main Stream of Music and Other Essays*, by Donald Francis Tovey, 271–97. Oxford: Oxford University Press, 1949.
- . “Sonata Forms.” 1929. In *Musical Articles from the Encyclopaedia Britannica*, 208–32. London: Oxford University Press, 1944. Reissued in 1956 under the title *The Forms of Music*.
- Webster, James. *Haydn’s “Farewell” Symphony and the Idea of Classical Style: Through-Composition and Cyclic Integration in His Instrumental Music*. Cambridge: Cambridge University Press, 1991.
- . “Sonata Form.” In *The New Grove Dictionary of Music and Musicians*, 2d ed., edited by Stanley Sadie and John Tyrell, 23:687–701. London: Macmillan, 2001.

Abstract

This essay provides a concept-history, a close examination, and a testing of the much invoked “sonata principle” (Edward T. Cone, 1967–68), while also introducing a new, contrasting mode of analysis (“Sonata Theory”) for sonata-form compositions from the decades around 1800. Within these compositions it was a common expectation that (nontonic) secondary- and closing-material from the exposition would normally be restated in the tonic in the corresponding place in the recapitulation. In mid- and late-twentieth-century English-language analysis (in response to shifting analytical paradigms), this expectation came to be inflated into differing recastings of a much freer, more encompassing “sonata principle” that, at least initially, was proposed to be the “unifying [and] underlying . . . principle for the Classical style.” Some of the attractions of this idea included its claims to midcentury academic sophistication, its protean flexibility, and its ability to provide quick solutions to otherwise “difficult” moments within highly regarded compositions. Anticipated by the caveats of other writers, this article calls attention to the principle’s limitations and the ways in which it has been imprecisely laid out or misapplied in influential writing. In a few comparative analyses I also present aspects of a more hermeneutically productive mode of analytical questioning “beyond the sonata principle.”