

Of Beginnings and Ends

A Corpus-Based Inquiry into the Rise of the Recapitulation

Yoel Greenberg

Abstract This article investigates the sources of the recapitulation using statistical methods. The recapitulation has traditionally been viewed as an expansion of small ternary forms, resulting in a top-down approach, whereby the repeat of expositional material is explained in rotational terms. Here I present a bottom-up approach, demonstrating that the recapitulation arose as a concatenation between two previously independent practices: the double return of the opening theme in the tonic in the middle of the second half of a two-part form, and the thematic matching between the ends of the two halves of two-part form. Drawing on a corpus of more than seven hundred instrumental works dated 1650–1770, I demonstrate that these two practices arose and functioned independently from each other, increasing in frequency and in length, before being subsumed into an overarching rotational practice.

Keywords sonata form, big data, recapitulation, double return, end-rhyme

THE REVIVAL OF *Formenlehre* over the past decade and a half has set the study of musical form on a new footing, rehabilitating discarded approaches on renewed foundations based on newly found sensibilities and insights, and circumventing the kinds of pitfalls that brought about the rejection of these older theories in the first place. William E. Caplin's (1998) *Classical Form* attempts to establish traditional *Formenlehre* "on more secure and sophisticated foundations," whereas James Hepokoski and Warren Darcy's (2006) *Elements of Sonata Theory* draws consciously from recent theoretical developments in such diverse fields as reader-response theory and hermeneutics in

I thank the following for their help at the different stages of writing this article: Naphtali Wagner, Markus Neuwirth, Scott Burnham, and Shari Greenberg for their helpful comments on the manuscript and on the dissertation from which it stemmed; Yair Goldberg, Laurence Freedman, Shlomit Yaari, and Shmuel Even-Zohar for their advice on statistics; and, last but not least, my beloved wife, Shlomit Yuval-Greenberg, who has the added benefit of being a neuroscientist and hence a seasoned statistician. Any inaccuracies are, of course, my own responsibility. This research was supported by the Israel Science Foundation (grant no. 190/15).

their fashioning of a new and detailed theory of sonata form. An equally important theoretical advance that can greatly benefit the new *Formenlehre* is the possibility, here in the age of big data, to manipulate unprecedented amounts of data with newly available computational tools. With the exception of a few as yet isolated attempts (e.g., Byros 2009; Rabinovitch 2015),¹ heavy-duty corpus-based studies on eighteenth-century repertoire, such as those by Robert O. Gjerdingen (1988, 2007) and Dmitri Tymoczko (2011), have not dealt with *Formenlehre* proper. In addition, the recent flourishing of *Formenlehre* research has left a significant subfield behind, for neither Caplin nor Hepokoski and Darcy (and very few of the studies written in their wake) has devoted significant attention to the genesis, or the evolution, of sonata form.

In the present study, I propose to take a step in this direction, examining the rise of the recapitulation through a data-driven approach.² Besides attempting to answer a question that pertains to the heart of sonata theory,³ I hope to show how careful quantitative research can yield qualitative insights into questions of musical form.

But is not the rise of the recapitulation already sufficiently understood? Is it not simply an expansion of the small-scale rounded binary forms into a large-scale form appropriate to genres such as the symphony and the string quartet? Such an explanation has indeed been widely accepted for at least half a century. Taking their cue from Heinrich Christoph Koch's (1983 [1793]) gradual expansion of a simple melodic phrase into a fully fledged sonata exposition,⁴ writers such as Eugene K. Wolf (1981, 573–96), Jan LaRue (1992, 184), Charles Rosen (1997, 83–88), and Joel Lester (2001, 50) have viewed various sonata elements as expansions of smaller-scale elements at the phrase level.⁵ The recapitulation was thus assumed to be a carryover from simple or rounded binary three-part form. Here I challenge this approach, using statistical methods on a corpus consisting of works from the century preceding the heyday of sonata form to show that the recapitulation arose as the concatenation of two distinct practices: the double return of the tonic key and the

1 Perhaps unsurprisingly, neither of these studies is devoted exclusively to *Formenlehre*. Rather, they attempt to bridge between Gjerdingen's (1988, 2007) corpus-based studies of eighteenth-century schemata and current theories of sonata form.

2 This article is part of a larger corpus-based project examining the role of repetition schemata within the evolution of sonata form. Here I examine only repetitions incorporated in the recapitulation, and these only in sonata form types 1–3.

3 The recapitulation has been described as sonata form's "central structural event" (Webster 2001) or "one of the great formal solutions of the high classical style" (Denny 1988, 352).

4 See Koch 1983 [1793], 129–65, and Ratner 1956. Koch, in turn, took his cue from Joseph Riepel (1754). For an excellent discussion of this approach and its relevance to Haydn's practice, see Sisman 1982.

5 Designations have differed, including LaRue's pseudo-evolutionary "timeline," Eugene Wolf's "expansion at phrase level," Rosen's "expanded Classical phrase," and Lester's "recursive expansions," but the ideas are similar.

main theme in the middle of the second half of a two-part movement, and the “end-rhyme,” that is, thematically parallel endings of the two halves of a two-part movement.⁶

Early descriptions of sonata form recognized that the recapitulation tends to include these two major chunks of faithful (or at least faithfully transposed) repetition, separated by an area that must be modified vis-à-vis the exposition. For instance, Francesco Galeazzi, in the second volume of his 1796 *Elementi teorico-pratici di musica* (Churgin 1968, 195–97), proposes a three-stage model for the recapitulation consisting of the return of “the first Motive of Part I in the proper natural key in which it was originally written,” followed by an area of loose repetition (or no repetition at all), whereby “the motive itself be conducted gradually to the subdominant of the key.” Finally, we return to exact repetition of expositional material: “Repetition of the last three periods of the first part is made transposing them to the principal key, and writing them after each other, in the same order they had in the first part.” A similar description was offered by Koch (1983, 201 [1793, §103]), according to whom “the last period of our first allegro [i.e., the recapitulation] . . . begins with the theme in [the main] key,” after which “phrases are compressed, as it were,” before “the second half of the first period . . . is repeated in the main key, and with this the allegro ends.” During the modified section, both Koch and Galeazzi recommend making a feint toward the subdominant, which necessarily involves compositional changes with respect to expositional material. The recapitulation thus consists of repetition of material from the beginning of the exposition, followed by freer material, and rounded off by repetition of the material that closes the exposition.

Nevertheless, we normally consider these two blocks of repetition as deriving from a single, overarching rotational logic. The appearance of a double return engenders expectations of an entire rotation—that is, the principal theme plus the “modified section” and the second half of the second half of the exposition, now transposed to the tonic—opening up the wealth of possibilities that enable dialogic form. Such expectations are the basis of Hepokoski’s discussion of the truncated recapitulation—that is, a recapitulation that does not state the second theme: “A sonata abandoned three-quarters of the way through, an incomplete sonata or sonata ‘fragment’ . . . a recapitulation lacking its second half, that very portion that . . . is supposed to define what we imagine sonata form to be” (Caplin, Hepokoski, and Webster 2010, 76). For Hepokoski, the truncated recapitulation is a “structural deformation” because it fails to fulfill the expectations created by the movement’s otherwise clear sonata-like signals (86). By the time Hepokoski’s two chief examples—Mozart’s overture to *Idomeneo* and Beethoven’s *Die Ruinen*

⁶ The term *end-rhyme* was coined by Leonard Ratner (1980, 212) in conjunction with both “small” and “large” binary forms.

von Athen—were composed, sonata form was a firmly established norm. Hepokoski's view of the forms of these overtures as being in dialogue with sonata form is therefore both a plausible and productive frame for interpretation.⁷ Yet, as with any dialogic form, we cannot escape asking: From what historical moment is the notion of dialogic form plausible? From what time on can we assume that listeners had enough information to form an educated set of expectations about how a musical work should behave?⁸

This is not merely an academic rumination: double returns followed by nonrotational material have always been a viable compositional option. Churgin (1984, xxvi) described such a precursor of the truncated reprise, frequently found in G. B. Sammartini's works from the 1720s to 1740s, as "a type of sonata form with incomplete recapitulation." Other "incomplete recapitulations" appear in the works of Sammartini's contemporary Antonio Brioschi, in whose symphonies "the musical material coming after the restatement of the primary theme or themes is often reformulated" (Mandel-Yehuda 1999, 2). Terms or words such as *incomplete recapitulation* and *reformulated* pre-empt Hepokoski and Darcy's notions of "writing over," implying an underlying ideal model that these movements fail to complete.⁹ In terms of listeners' expectations, these terms suggest that, given the double return, we would now expect a complete recapitulation, and that any new material from here on must be considered as a writing over of expositional material. The finale of Sammartini's Symphony no. 1 J-C 7 (Example 1) is a typical example of such a form. Measures 1–10 are restated in a slightly simplified version in mm. 47–54, but the remaining twenty-two measures have little more than a vague thematic relationship to the material in the exposition, and certainly nothing of the kind associated with a sonata recapitulation. With this and other examples dated by Churgin around 1730 (they were published in a collection from 1741), and with numerous other contemporary examples by Brioschi, Chelleri, Somis, and even J. S. Bach, we cannot but ask ourselves what there is to be gained by describing this as "sonata form with incomplete recapitulation." In terms of dialogic form, could listeners possibly have encountered enough fully fledged recapitulations for their expectations to have been thwarted when one post-double-return space does not make the grade? The problem becomes even more pressing if we take into account the acquaintance of an audience with one particular composer: when we are told that Brioschi and

⁷ Caplin has taken issue with the details of Hepokoski's analysis (Caplin, Hepokoski, and Webster 2010, 92–93). What is pertinent here is not the analysis, however, but the analytic option of viewing the truncated recapitulation as a deformation of sonata form.

⁸ The notion of the false recapitulation, perhaps the oldest standing dialogic idea vis-à-vis sonata form (although it predates Hepokoski by many years), has been challenged by Peter Hoyt (1999) and Markus Neuwirth (2013) on similar chronological grounds.

⁹ As Hepokoski and Darcy (2006, 214) write: "One theme writing over a space normatively occupied by another is a central aspect of Sonata Theory. This approach attends not only to what happens acoustically in a piece—what we hear—but also to the things that we expect to happen that do not occur or that are kept from sounding."

Presto

11

... “Development section” (26 measures) ...

47

Double return

55

Non-expositional material

p

61

tr

VI. I, II

69

VI. I, II

Example 1. G. B. Sammartini, Symphony in C major J-C 7, III Presto, mm. 1–20 and 47–76

Sammartini in their pre-1740 symphonies strongly tend toward reformulation of the area following the double return, surely their listeners could no longer have expected a recapitulatory rotation solely on the basis of hearing a double return. Instead, as suggested by Markus Neuwirth (2013, 285), the double return would have functioned as a local thematic marking of the

return to the home key, without any rotational significance, providing the listener with a sense of orientation but without creating any definite thematic expectations (see also Hoyt 1999, 27).

Churgin's classification of these cases as "incomplete recapitulations" shares with the phrase-expansion hierarchical approach the assumption that the governing principle of the recapitulation is one of large-scale repetition of sequences of material (in what will eventually provide a basis for Hepokoski and Darcy's notion of "rotational logic"), and that the locally repeated modules are derived from a more general principle of repetition. This approach may be classified as "top-down" in that it explains lower-level individual parts as governed by a rationale at higher levels. But in light of the numerous examples in the period under question of double returns without end-rhymes, or of end-rhymes without double returns,¹⁰ it might make more sense to reverse the arrow of causality. Rather than assuming that the double return was the result of the wish to include a recapitulatory rotation, we may postulate that the recapitulatory rotation took advantage of separate, preexisting building blocks such as the double return and the end-rhyme and combined them into one, overarching entity. The resultant model would thus be "bottom-up," with the high-level entities emerging on the basis of already present lower-level practices. Put in terms of listener's expectations, had the double return always engendered expectations of an end-rhyme, or was there a time when the presence (or lack of presence) of the double return would not have enabled listeners to create any end-rhyme-related expectations whatsoever?¹¹

Listeners learn to expect what is frequently encountered, and their expectations are thwarted upon the occurrence of a rare event. Frequency and rarity, and hence expectations, are ideally measured through statistics, so I investigated this question through a statistical study, analyzing the mutual dependence of the double return and the end-rhyme in a corpus of instrumental works predating the rise of sonata form. Questions of dependence, sometimes referred to as second-level statistics (as opposed to first-level statistics, which are restricted to simple questions of incidence of phenomena, rather than relationships between them), are only rarely employed in music research, especially in research of musical form.¹² The statistics involved are

10 We are more accustomed to end-rhymes without double returns, as they are characteristic of typical baroque binary form.

11 Asking different questions and applying different methods, Neuwirth (2013) reaches conclusions very similar to my own regarding the initial role of the double return.

12 Gjerdingen's work is a welcome exception, but it does not engage with *Formenlehre*. Indeed, it views the notion of sonata form as "more a hindrance than a help in understanding how galant compositions were made and understood" (Gjerdingen 2007, 22). Vasili Byros (2009, 2015)

and Gilad Rabinovitch (2015) have attempted to reconcile Gjerdingen's approach with *Formenlehre*, both making significant use of statistics. Also of interest to the relevant repertoire, yet not directly concerned with form, is Uri Bin-yamin Rom's (2011) dissertation on Mozart's use of tonality. In earlier research, most statistical corpus studies were limited to the music of single composers. A remarkable exception is the pioneering research of Roger Kamien (1966). Kamien's method of random sampling (a necessary limitation in the pre-personal computer era) excludes the possibility of extracting any second-level data due to the limited size of the sample.

not complicated, but since many music scholars may be unfamiliar with them, I include a brief definition of statistical dependence/independence:

Two variables A and B are statistically independent (or, simply put, independent) if the incidence of encountering both together is equal to the product of the incidences of both variables, that is, if $P(A \cap B) = P(A) \times P(B)$.

More intuitively, independence reflects whether or not the occurrence of one event provides us with any new information on the probability of the other. It is therefore, by definition, the ideal tool for examining the question in hand: whether or not, given a double return, we should entertain increased expectation of an end-rhyme.

In this case, the two variables are the double return, that is, the tonic repeat of the thematic material from the opening in the middle of the second half of an instrumental binary movement, and the end-rhyme, or a significant thematic matching between the endings of both halves of an instrumental binary movement.¹³ If the number of works with both a double return and an end-rhyme is roughly the same as the product of the frequencies of both those features, one can conclude that the double return and end-rhyme are independent. If, however, the number of works with both features is significantly more than the product of their frequencies, one can conclude that they are dependent and that the appearance of one indicates the appearance of the other is more likely. If the number of works with both features is significantly less, one can conclude that they are dependent but that the appearance of one indicates the appearance of the other is less likely. In short, to determine whether or not the appearance of a double return would justify increased expectations of an end-rhyme,¹⁴ I (1) measured the incidence of double returns ($P[A]$), (2) measured the incidence of end-rhymes ($P[B]$), (3) measured the incidence of double returns followed by end-rhymes ($P[A \cap B]$), and (4) checked whether the condition for independence, $P(A \cap B) = P(A) \times P(B)$, holds. We know that eventually (certainly by the end of the eighteenth century, when Koch and Galeazzi wrote their descriptions of sonata form, but probably considerably earlier too)¹⁵ composers had a pretty good

13 It is necessary to well define what qualifies as a double return and what as an end-rhyme: How much repetition is needed for the return to be significant? How exact must the restatement be? I address these questions after laying out the procedure for analyzing the sample repertoire. Both these features have been explored statistically in earlier studies limited to the works of single composers. Michael Talbot (1996) examined their presence in Arcangelo Corelli's works to prove that Corelli was more "up-to-date" (from a formal point of view) than was previously acknowledged; James Webster (1982) examined their presence in early works by Joseph Haydn. Both articles were limited to first-level statistics (questions of frequency), although Talbot (1996, 156) attempted to draw a conclusion about the relationship between the two ele-

ments, writing that Corelli tends to "ration overt requotation as if he regarded it as too obvious or facile." Unfortunately, this was not shown rigorously; in fact, his data suggest the very opposite.

14 The existence of both a double return and an end-rhyme is a necessary but not sufficient condition for a recapitulatory rotation. I address connection to rotational logic below.

15 Koch scholarship maintains that Koch's description reflects the practice of the 1760s and 1770s rather than the period during which his treatise was written. As I argue in the last section of this article, this is not entirely true. Koch's description is indeed based on repertoire composed

idea of what a recapitulation was, so I conducted the test for independence diachronically, testing for every decade within the sample whether or not independence holds.

The sample

The definitions and procedure outlined above suggest the following six restrictions on the sampled repertoire:

1. The diachronically sensitive method allows examination only of works with a date, at least approximate.¹⁶
2. The definitions of the double return and the end-rhyme restrict the sample to two-part movements alone, and hence only movements with binary repeat signs were included. Movements with obvious two-part formal logic but without repeat signs were not considered, as they would add an unwanted degree of subjectivity (although not unheard of, these are relatively uncommon in the period under examination and would have been unlikely to skew the sample in any way).

It is important to emphasize that, due to this restriction, the very selection of the sample represents a departure from the approaches of Rosen and LaRue, both of whom argued for an influence on sonata form of nonbinary forms such as ritornello and da capo forms. To support my decision, I cite the fact that sonata form was considered a subspecies of binary form until well into the nineteenth century (e.g., the quote from Johann Friedrich Reichardt in the “Discussion” section of this article, Antoine Reicha’s *grande coupe binaire*, or even Beethoven’s own manuscript markings of *prima* and *seconda parte* [Drabkin 1991]), so examining audience expectations and compositional choices within this category is historically justifiable. Second, and more important, as will presently become apparent, the period I am examining is contemporary with, not later than, the period of ritornello and da capo forms. The evolution of those forms therefore runs parallel to similar evolution within binary forms, and there is thus no more reason to assume that it was the influence of ritornello form that brought on the rise of the double return in binary forms than there is reason to assume that ritornello form inherited the double return from early binary forms. Instead of assuming that one form influenced the other, it appears likely that forms with double returns evolved inde-

twenty to thirty years earlier, but his interpretation of that repertoire is not necessarily in concordance with the way the composers of those works would have viewed them.

¹⁶ Dates of composition have been taken from work lists in the *Grove Dictionary of Music and Musicians*, from thematic catalogs, or from the editions themselves. Where no information on the date of composition has been available, date of publication has been taken, unless it is clear that

the work was published long after its composition. Works whose year of composition is not accurately known have been dated according to the latest possible date (e.g., –1728 becomes 1728) and works in which a range of years is given have been dated as an average of the extremes (e.g., 1726–30 becomes 1728). Where no date at all was available, the work was not included in the sample.

pendently of each other as the result of what James Webster (1982, 127) has described as “the growth of a feeling for the double return.”

3. Only pieces for chamber ensembles or for solo keyboard are included, with an eye to reducing problems of orchestration that could complicate decisions as to what may be considered a repeat. Leaving orchestral works for further study is also consistent with the eighteenth-century priority for distinction between genres over distinction between forms.
4. I have limited the sample to chief centers of influence in the development of instrumental music in the early eighteenth century (Italy, North Germany, South Germany, and Austria) to try to reconstruct a “mainstream” narrative of the evolution of sonata form. This leaves out a number of other important musical centers, such as France, which contributed little to the repertoire of instrumental music in the second half of the eighteenth century most closely identified with sonata form. The most significant result of this limitation is that the keyboard sonatas of Domenico Scarlatti were not included, having been composed outside the examined spheres of influence (those of them that may have been composed in Italy pose significant problems regarding their chronology).¹⁷
5. Only movements in a moderate to fast tempo are included.¹⁸ It has often been noted in sonata-form literature that a full-blown recapitulation is more characteristic of moderate to fast-paced movements than of slow movements. This indicates that double returns behave differently in slow movements than in faster ones, suggesting that slow movements should be examined separately.
6. The sample is selected so as to be representative, at least as closely as possible, to the overall creative environment in the relevant period.

Obviously, an entirely accurate reconstruction of the period is impossible, but this should not deter us from attempting to balance the sample in the best way possible. I attempted to take various parameters into consideration, including the quality of the composers (all other things held constant, composers accepted as great masters can be assumed to have been more influen-

17 “Influence” is a Janus-faced concept, looking both backward and forward. In the case of Scarlatti, Haydn’s epithet that working in geographic isolation “forced” him “to be original” seems to hold even truer than for Haydn himself. Yet these works were nevertheless influential, even if not contemporary to their composition. See Sutcliffe 2003. Given a reliable chronology for Scarlatti’s sonatas, it would be interesting to compare their evolution with that of binary sonatas toward sonata form in the area of interest of this study. The comparison would be akin to the

way evolutionary biologists compare evolution of species on the mainland to those on “evolutionary islands.” Unfortunately, chronologies of Scarlatti are subject to intensive debate and are thus too speculative to allow for such a research.

18 I also include works where a fast or moderate tempo was obvious from the musical material, even if not indicated explicitly.

tial than *Kleinmeister*), their fecundity (prolific composers would have more influence than those less so), and their geographic location (important cultural centers vs. peripheral courts). Specific historical considerations were also taken into account, as with Arcangelo Corelli, who was considerably more influential than he was prolific (Allsop 1999, 175–99). Likewise, published works would have been more influential than unpublished ones, and works with many editions more than those printed only once. Fortunately, processes of canonization and preservation reflect much of this, if not entirely accurately, then at least closely enough for present purposes. Scanning two or three good libraries was therefore a good place to start (in this case the Mendel Music Library at Princeton; the music library at the Hebrew University, Mt. Scopus, Jerusalem; and the online Petrucci Music Library at imslp.com). Some corrections had to be made to ensure that highly prolific composers with “complete editions” (e.g., Vivaldi and Telemann) were not over-represented, either in the sample as a whole or within every decade in it. The sample is therefore not entirely random but artificially constructed to reflect contemporary practice as closely as possible. The final sample includes 732 fast or moderately paced binary instrumental works dating 1650–1769 (grouped by decade) by eighty-four composers born and active in Italy, North and South Germany, and Austria. Due to the paucity of exemplars before 1700, I have taken 1650–89 as one “long decade” (marked “–1689”) in all first level data, and 1650–99 as a “long decade” in all second-level data (marked “–1699”). A complete list of movements included in the sample can be found at yoelgreenberg.wixsite.com/beginnings-and-ends.

Gathering the data

For each work I have determined whether it features a double return and whether it includes an end-rhyme. In the overwhelming majority of cases, determining the answer was straightforward. There were, however, a number of problematic, yet typical, musical situations that had to be considered systematically to assure consistency. From a thematic point of view, the greatest challenge was to determine the limits of what should be considered a repetition, and here I chose different solutions for each of the examined features.

In the case of the double return, three typical problems presented themselves, the first two thematic and the third harmonic: paraphrased repeats, shortened repeats that pick up from the middle of primary theme (P) material, and extratonic reprises (most commonly subdominant, less frequently submediant).¹⁹ Unlike off-tonic references to opening material at the

¹⁹ Besides the better-known instances in Schubert’s music, Florian Leopold Gassmann (1729–74) had a penchant for subdominant and submediant reprises (also in major keys), and from the 1780s Clementi experimented with reprises in several off-tonic keys.

double bar, which, as observed already by Koch (1983, 236 [1793, §151]), are frequently transformed through inversion (e.g., J. S. Bach, Allemande from the Partita in D minor for solo violin) or limited to rhythmic similarity (e.g., the Allegro from Corelli's Trio Sonata op. 2/2),²⁰ paraphrased repeats are relatively uncommon in double returns. I ran separate analyses of the data, first counting and then discounting paraphrased repeats as double returns, with essentially identical results (*vis-à-vis* independence). In the data presented here I counted paraphrases as repeats. Shortened repeats—double returns that occur in mid-P material—were not counted as double returns unless they occur on a clearly recognizable variant of the opening music (such as a double return that picks up from the A' in a movement beginning with an AA' or ABA' structured P section, e.g., the first movement of Haydn's op. 9/1). This is because the main impact of the double return is local and immediate. Repeats of P that do not achieve such an impact cannot, therefore, be considered true double returns.

Off-tonic reprises present a more difficult decision. On the one hand, an off-tonic reprise functions in many cases exactly like a tonic reprise, provided it is harmonically stable; on the other hand, many works include stable references to the main theme in nontonic keys between the double bar and the return to the tonic. In a work lacking a double return, deciding whether a subdominant reprise counts as a substitute for a double return would be speculative and subjective.²¹ I have therefore excluded works with a nontonic return from the sample, and while this may affect the overall result, it is rare enough an occurrence to make any such effect marginal.

A more frequent but equally problematic case is that of the mid-eighteenth-century stereotype dubbed the "premature reprise" by Oliver W. Strunk (1974, 126–70) and Rosen (1988, 155–56): the repetition of the main theme in the dominant at the double bar and then immediately afterward in the tonic—an option already described by Koch (1983 [1793], 235).²² I have not counted this as a valid double return, because it does not function as a marker of the final return to the tonic. Instead, it acts as a foil, helping to underscore the nature of the dominant repeat of the opening theme at the double bar as an antithesis to the original theme, thus highlighting the harmonic difference by means of close juxtaposition.²³

20 Koch's example for a modified return is his own. His example for an inverted repeat is from the first of C. P. E. Bach's Prussian sonatas.

21 For a discussion of a particularly ambiguous case in Haydn's op. 2/4, see Rosen 1988, 152.

22 The premature reprise has also been referred to as the "immediate reprise" by Webster (1982, 128), the "precursory recapitulation" by Mark Evan Bonds (1988, 225), the "medial return" by Hoyt (1999, 7), and the "early medial return" by Neuwirth (2013). The multiplicity of names

reflects the challenges presented by this practice to theories of sonata form.

23 Although in some cases (e.g., Haydn's op. 1/2), there eventually turns out to be no other double return (what Bonds has named the "disjunct" recapitulation), the moment of return is nevertheless only belatedly interpreted as such. It would perhaps be better to view these cases as lacking a true double return, with the omission justified by the premature reprise.

The case of the end-rhyme is somewhat different. The end-rhyme consists of music from the end of the first half, and unlike the opening theme, the nature of the thematic material involved is generally less arresting. Where the double return invokes opening material, which according to Galeazzi must be “well rounded and lucid” (Churgin 1968, 191) and hence easily recognizable, the end-rhyme frequently picks up from the first half in the middle of sequential passagework and hence becomes apparent only with the passage of time. This is probably why the end-rhyme is even less prone to significant thematic transformations than the double return. One type of transformation to which it may be subject is variation in the number of sequential units within passagework, a result of the difference between the harmonic trajectories of the two halves. This, however, does not diminish the thematic parallelism between the halves and thus does not present a problem. The indistinct thematic profile of the end-rhyme makes its length an important factor, granting the listener more time to become aware of the parallel, or providing more information for the connection to be made. Therefore one would expect different results for different thresholds for the length of the rhyme. Here, too, I ran statistical analyses on various thresholds. The results regarding independence of the two types of repeats were invariable under change of threshold.

These nontrivial decisions notwithstanding, in the overwhelming majority of cases determining the presence or absence of a double return or an end-rhyme was clear-cut, and due to the large sample, any such decisions are unlikely to have significantly altered the results or to have introduced any systematic bias.

Results

First-order results of the data on the sample are shown in Figure 1. As one would expect, the frequencies of both the end-rhyme and the double return increase significantly over the sampled period. The end-rhyme was fairly common at the beginning, appearing in approximately one of every three cases, while the double return was a rarity. This difference between the two can be explained by the relatively short lengths of instrumental movements in the late seventeenth century: an end-rhyme appears at the end of the second half, and every movement, however short, has an end; a double return appears in the middle of the second half, and a short movement does not necessarily have enough space in its second half to introduce a significant thematic quote in its middle.²⁴ Put differently, the tripartite division suggested by the double return may require greater length than the bipartite one sug-

²⁴ Of course, the causality could work the other way round: what Webster (1982, 127) has described as “the growth of a feeling for the double return” could have led to an increase in movement length.

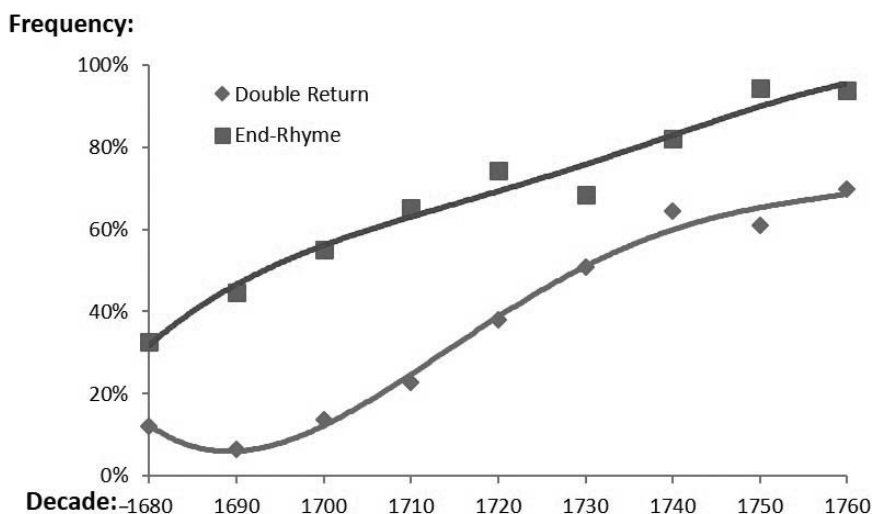


Figure 1. Double returns and end-rhymes, 1650–1760. Data points represent measured data; trend lines are for illustration only.

gested by the binary repeat sign (and endorsed by the end-rhyme). By the end of the period covered by the sample, the end-rhyme has become *sine qua non* and the double return seems well on the way there too, a likely scenario where the end point marks the beginning of the heyday of sonata form, which was characterized by both kinds of repetition. The slight dip in frequency of the end-rhyme during the 1720s is probably an insignificant artifact of the sample (i.e., different works would probably have shown the same trend but without the same dip).

Figure 1, although interesting in its own right, does not shed much light on the formation of the recapitulation. It is certainly compatible with my own hypothesis—that the double return and end-rhyme arose as independent phenomena—but it is also explained by LaRue’s (1992, 188–89) timeline, whereby binary forms characterized by end-rhymes but not by double returns dominate the early phase, gradually giving way to sonata forms characterized by both. What is essential to decide between the two approaches is the question of dependence between the variables.²⁵ As explained earlier, according to the definition of statistical independence, the double return (DR) and end-rhyme (ER) are independent if $P(\text{DR} \cap \text{ER}) = P(\text{DR}) \times P(\text{ER})$. Therefore, under the assumption of independence (i.e., a bottom-up approach), we would expect to see a proportion of $P(\text{DR}) \times P(\text{ER})$ in works that have both a double

²⁵ The following few pages include a discussion that although not complicated, may be palatable mainly for the mathematically inclined. Readers may wish to skip the rest of the “Results” section and pick up again at “Discussion.”

return and an end-rhyme. We can therefore write $P_{BU}(DR \cap ER) = P(DR) \times P(ER)$, with P_{BU} denoting the proportion of works with both double returns and end-rhymes predicted under my bottom-up hypothesis.

P_{BU} must be compared with $P_{OB}(DR \cap ER)$, that is, the observed proportion of works with both a double return and an end-rhyme within the sample. If the two values are close, this will demonstrate independence, and if not, dependence. What may be considered “close”? There is no definitive answer to this, but we can offer a comparative scale by asking what would be expected under a top-down model, which I denote $P_{TD}(DR \cap ER)$. A top-down model would assume that, at least ideally, every double return would have a corresponding end-rhyme. Allowing for the occasional “incomplete reprise” one could probably give or take a few cases, but overall, one would expect something close. Therefore, under a top-down model, the maximal possible value for $P_{TD}(DR \cap ER)$ would be the lower of the two values, $P(DR)$ and $P(ER)$. The rationale is that having both a double return and an end-rhyme is a special case of having a double return; therefore, $P_{TD}(DR \cap ER)$ cannot be greater than $P(DR)$. It is also a special case of having an end-rhyme, and therefore it cannot be greater than $P(ER)$. The resulting pair of inequalities $P_{TD}(DR \cap ER) \leq P(DR)$ and $P_{TD}(DR \cap ER) \leq P(ER)$ leads to $P_{TD}(DR \cap ER) \leq \min[P(DR), P(ER)]$.²⁶ In sum, I compared three values:

- The observed incidence of works with both elements: $P_{OB}(DR \cap ER)$
- The expected incidence under a bottom-up model: $P_{BU}(DR \cap ER) = P(DR) \times P(ER)$
- The expected incidence under a top-down model: $P_{TD}(DR \cap ER) = \min[P(DR), P(ER)]$

If P_{OB} is significantly closer to P_{BU} than to P_{TD} , I will have proven my bottom-up hypothesis.

Figure 2 shows the results of this comparison over my sample. The columns representing P_{BU} (hatched) and P_{OB} (solid) are consistently close to each other, whereas P_{TD} is consistently much higher. Furthermore, values for P_{BU} are consistently very close to those for P_{OB} , sometimes a little higher and sometimes a little lower (a variance that suggests that these differences should not be overinterpreted). Thus P_{BU} predicts P_{OB} extremely well overall.

Another indication of the dependence between the double return and the end-rhyme may be obtained by calculating the correlation between them. Correlation represents the relationship between two measured data values through the correlation coefficient, which ranges from -1 to $+1$, where -1 indicates perfect negative correlation, 0 indicates no correlation, and $+1$ indicates perfect positive correlation. In other words, a correlation of 1 between

²⁶ Specifically for my case, as Figure 1 shows, the incidence of the double return is lower than that of the end-rhyme across the entire time span, so the expected value of $P_{TD}(DR \cap ER)$ would be a little less (allowing for the

occasional deformation) than the incidence of the double return. Therefore, for the top-down model, in theory $P_{TD}(DR \cap ER) = \min[P(DR), P(ER)]$, and in practice this will end up equaling $P(DR)$.

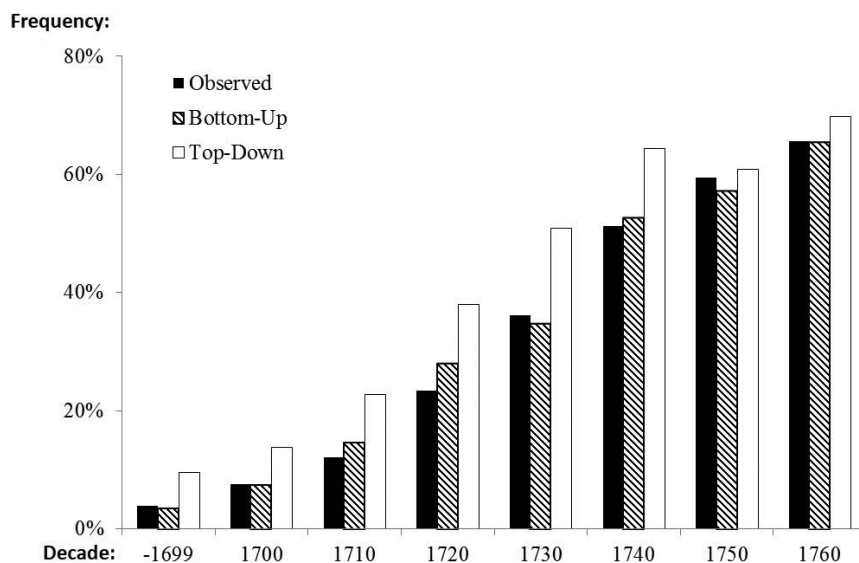


Figure 2. Intersection of double return and end-rhyme, 1650–1760. Black column, P_{OB} ; hatched column, P_{BU} ; white column, P_{TD} .

element A and element B means that an increase in A predicts with certainty a corresponding increase in B (but the increases would not necessarily be of the same scale). Similarly, a correlation of -1 means that an increase in A predicts with certainty a decrease in B (again, of unknown size). A correlation of 0 would mean very little dependence between the elements.²⁷ In our case, a near-zero correlation between the double return and the end-rhyme would therefore support the bottom-up hypothesis.

The results show that the correlation between the double return and the end-rhyme over the period of the sample (Figure 3) is, generally speaking, very low, beginning around zero (no connection at all) in the late seventeenth century, dropping to a negative correlation (as low as -0.27) in the 1720s (interestingly, this is the decade in which Churgin’s “incomplete recaptulations” were composed), and then climbing erratically to a low positive correlation by 1770.

The low correlations in Figure 3, which zigzag erratically around 0 , suggest that there is no correlation at all between the two elements. Moreover, none of the correlation coefficients are statistically significant, meaning that there is very little point in trying to infer anything from these data (other than the fact that this is what a zero correlation might look like). A list of cor-

²⁷ Here I use G. Udny Yule’s (1912) phi coefficient of correlation, which is equivalent to the most common measure of correlation—the Pearson correlation coefficient—but better suited to the binary values of my data (i.e., each ele-

ment has only two values, yes or no, unlike continuous data such as height, where the Pearson coefficient would have been more suitable).

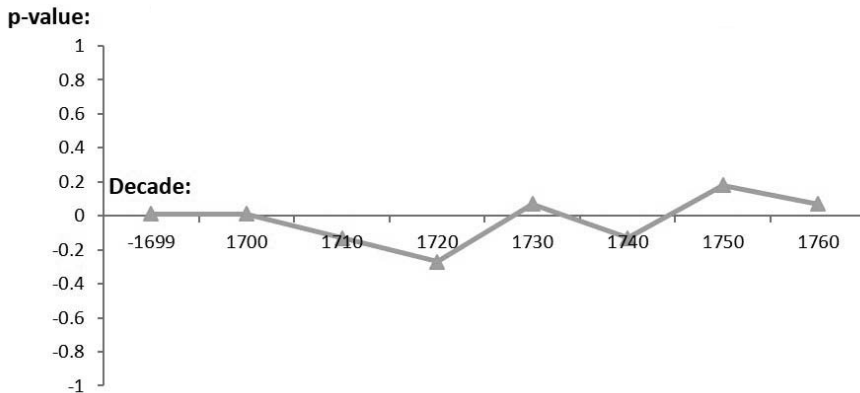


Figure 3. Correlation of double return and end-rhyme, 1650–1760

Table 1. Correlations per decade between the double return and the end-rhyme

Decade	Correlation (ϕ)	Significance (r)
1650s–1690s	0.012726	0.897481
1700s	-0.00365	0.97438
1710s	-0.13451	0.281572
1720s	-0.22303	0.012779 ^a
1730s	0.052434	0.589927
1740s	-0.08804	0.425827
1750s	0.182323	0.133764
1760s	0.017573	0.865062

^aThe negative correlation for the 1720s is statistically significant. However, wherever multiple comparisons are being made (as with the numerous decades examined here), a correction for multiple comparisons must be applied to the r values, rendering this value insignificant. In any case, a negative correlation is even worse than a zero correlation for a top-down approach.

relation coefficients and their corresponding significance measures (r values) is presented in Table 1.²⁸ The lack of significance might be due to a small number of works within each time frame. I therefore increased the time frames to include four decades each, creating a three-decade overlap between time frames (Table 2), with sample sizes of at least 250 works per time frame.

28 As is accepted in social sciences, I have taken a threshold of 0.05 for significance (i.e., the r value must be less than 0.05 for corresponding correlation to be considered significant).

Table 2. Correlations between the double return and the end-rhyme for overlapping forty-year time spans

<i>Time Span</i>	<i>Correlation (ϕ)</i>	<i>Significance (r)</i>
1650–1720	–0.00489	0.938541
1690–1730	–0.02985	0.596453
1700–1740	–0.04363	0.397617
1710–50	–0.0644	0.209166
1720–60	–0.02864	0.575324
1730–70	0.0623	0.240339
Entire Sample	0.124404	0.000743

As Table 2 shows, the correlations became even smaller, still with no significance. The last row in Table 2 shows the correlation between the double return and the end-rhyme over the entire sample. Because of the increase in sample size, the value is now statistically significant, but its value is very low: 0.124. This means that the percentage of explained variance between the double return and the end-rhyme over the entire sample is 1.5 percent, which is tantamount to zero.²⁹ In sum, the data strongly support my hypothesis that the double return and the end-rhyme are statistically independent, and the appearance of one would not have created among listeners any expectations to encounter the other.

Discussion

At this point I would like to raise several points. First, if the double return and the end-rhyme are independent of each other, why did they increase in frequency at the same time? As a matter of fact, why would either of them increase in frequency at all? The answer, I believe, lies in the combination of a variety of factors, but perhaps the most important of these is the increase in length of musical movements throughout the period in question. As is evident from Figure 4, two-part movements grew in length over the sampled period, starting as typically short pieces, including a brief first half of eight to sixteen measures and a second half of roughly the same length, and ending up as complex movements averaging ninety measures in length and including second halves substantially longer than the first halves. Such an increase in length, and hence complexity, necessitated the use of mnemonic devices, in particular repetition, to facilitate comprehension of the works and to maintain long-term coherence through recognizable thematic relationships. A more speculative explanation for the increase in popularity in repetition schemata is that the rising middle-class audience would have needed such devices more than the more sophisticated and better-educated upper-class audience they were gradually superseding. Either way, the increasing popu-

²⁹ The percentage of explained variance is obtained from the square of the correlation coefficient.

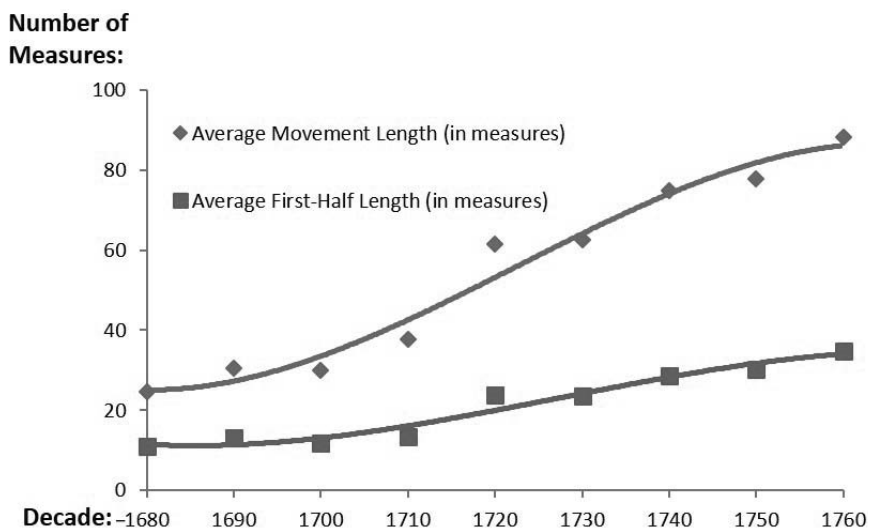


Figure 4. Increase in length of movements, 1650–1770

larities of the two repetition schemata were therefore the results of the same causes but did not depend on each other.

Second, my approach to form, and hence to the compositional process in general, may seem to run contrary to implicit notions of how creativity works. Surely the agency of the composer, who seeks to create a coherent work, will forge a connection between formal elements within a single work. As an otherwise sympathetic anonymous reviewer of the grant that funds this research wondered: “Any musical structure, be it a phrase, a period, a paragraph, a reprise, or an entire movement, is presumably informed by some overarching idea of its place within the larger whole. To suggest that a composition consists of essentially modular and wholly interchangeable elements runs counter to the evidence we see in any competently composed work, regardless of its form” (pers. comm.).

This is doubtless true, yet it is no more significant a problem than the use of statistically common linguistic constructs (e.g., grammar, vocabulary, and word combinations) by great authors and poets, or even by ourselves in what is presumably spontaneous speech. Composers inherit practices from their predecessors. As individuals within a community, they are susceptible to influences from within it; as communicators who share an audience with other communicators, they are subordinate to a shared set of unwritten rules. Craftsmanship and creativity commence only after their building blocks have been decided. One of the broader implications of my bottom-up approach is that the creative challenge, at least in the formal sense, results from the problem-solving nature inherent in the combination of independent, and not always easily compatible, preexisting practices. In other words, a composer

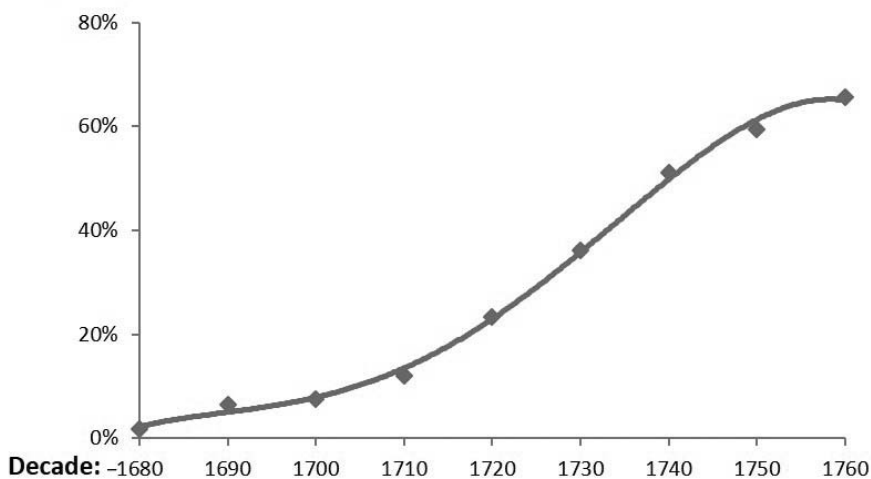
Frequency:

Figure 5. Co-occurrence of the double return and the end-rhyme, 1650–1770

may, and frequently will, create dependency between formal elements *within a single work*. But the choice of elements available to that composer is suggested (statistically, not deterministically) by a network within which those elements are independent.

Third, does it follow that the recapitulation as an integrated formal unit never really existed? Was it only a product of an incorrect observation that identified a single, unified entity where two separate ones were, in fact, present? The answer, clearly, is no. By 1793 Koch had described an entire package that was to become known as sonata form, with a full-blown recapitulation rounding off his description. Koch's *Versuch*, like Galeazzi's *Elementi teorico-pratici di musica* and like the overwhelming majority of descriptions of sonata form written in its wake, was a tutorial for would-be composers, and thus the descriptive would become prescriptive, creating a strong dependency between the double return and the end-rhyme through the overarching concept of recapitulation. An acquaintance with the sonata-form repertoire of Haydn and Mozart shows that this was the case a good twenty to thirty years prior to Koch. How, then, was the step made from the use of the two types of repetition as separate, independent elements to their emergent conception as parts of the recapitulation?

It is precisely the kind of dependency-creating practices I describe above that help explain how the recapitulation finally emerged. Consider the frequency of the co-occurrence of the double return and end-rhyme. Until the end of the second decade of the eighteenth century, co-occurrence of the double return and the end-rhyme was rare. Only one in ten works would have included both elements, so even had it occurred to the occasional composer to bind the two elements together under the logic of a recapitulatory rota-

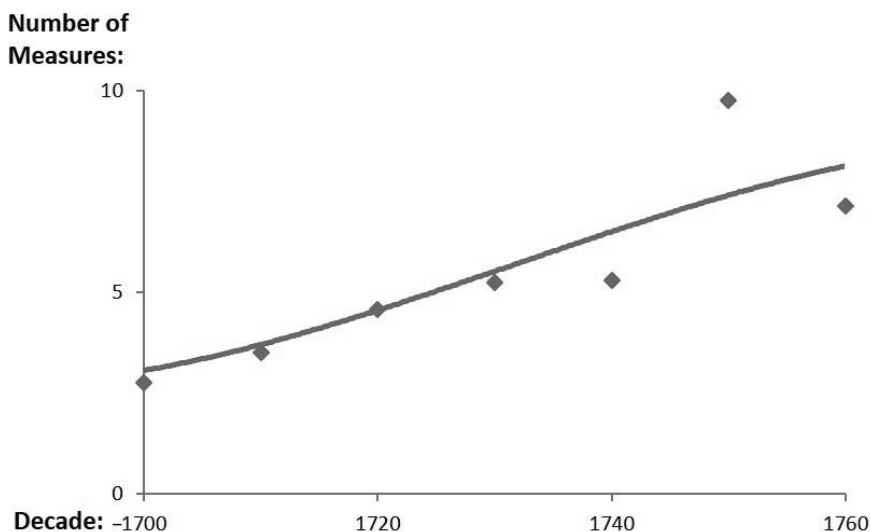


Figure 6. Average length of double return (in measures), 1650–1760

tion, this would have been no more than an isolated incident, unlikely to have had any long-term repercussions. Yet in the last three decades of the sample, roughly two out of every three two-part movements would have included both a double return and an end-rhyme. One can easily imagine how such a co-occurrence could have been interpreted as a single phenomenon comprising both elements rather than as two separate phenomena. A recapitulatory rotation would be one of the most obvious ways of rationalizing this misperception, providing a single framework that includes both elements.

Further support for this idea comes from the length of the two elements, that is, the extent of repeated material in the double return and the end-rhyme. Figure 6 presents the average length in measures of double returns across the sample (due to the paucity of double returns before 1710, data before 1710 have been averaged together). The double return started as a brief thematic statement, typically one to two measures long and thus consisting of little more than a brief thematic statement. It would have served as a very local thematic confirmation of the return to the tonic, enhancing our sense of having arrived back home. By the end of the period in the sample, the average double return was approximately eight measures long, thus consisting of a period-long restatement of opening material. Of course, this is in part the result of the increasing periodization of the musical style over the examined time span: in 1700 most works would not have started off with a detachable (and hence quotable) period, whereas by 1760 most works would. Nevertheless, an extended and frequently tonally complete restatement was more likely than a mere thematic snippet to have suggested an extension to a full rotation.

**Length (proportional
to first half):**

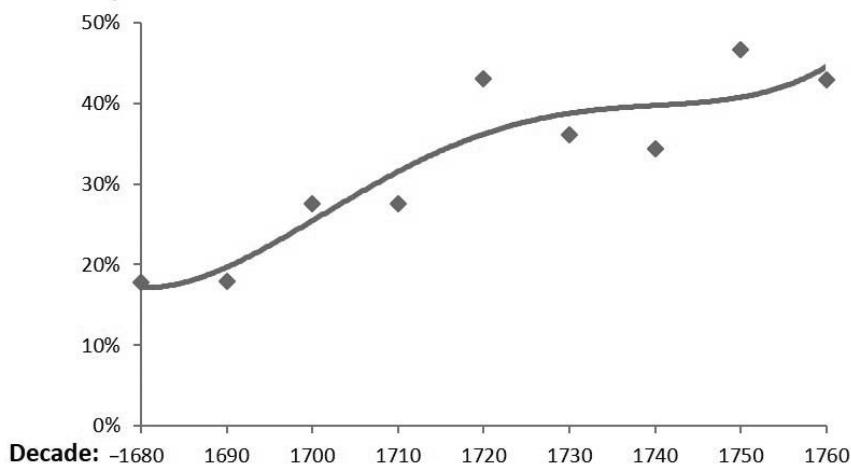


Figure 7. Length of end-rhyme (relative to the length of the first half), 1650–1760

In the case of the end-rhyme, this trend is even clearer. Figure 7 presents the length of the end-rhyme in terms of the percentage of repeated thematic material from the first half.³⁰ The earliest examples average less than 20 percent of the length of the first half. As a combination of the data in Figures 4 and 7 suggests, these typically take up one or two measures, consisting of little more than a brief cadential formula. Yet by the mid-eighteenth century nearly half of the expositional material is repeated, normally including at least a dozen measures. The combination of practices—a prominent restatement of the entire opening period, and the repeat of nearly all the second half of the exposition—would have been experienced as an entire rotational repetition even had the intention not been there, although it would have probably occurred to the composers too.

The double return and the end-rhyme thus started as local thematic markers, with the double return underscoring a significant tonal event and the end-rhyme confirming the two-part layout, but with virtually no connection between them, and set far apart from each other within the movement. In time, the increase in use of both these elements brought them to co-occur frequently across the repertory of binary works, and the increase in their length brought them within “touching distance” of each other inside single

30 As opposed to the double return, which relies on music with a clear thematic profile, the end-rhyme frequently commences surreptitiously, as part of sequential spinning out of nondistinct thematic material. Recognizing the parallelism with the first half can therefore take some time. Hence I have opted for different measures for each type:

the double return on the basis of absolute length (in measures), and the end-rhyme in terms of length relative to the entire first half. In light of the increase in movement length over the sample, an absolute measure of the length of the end-rhyme would have yielded even more radical results than those presented here.

movements. Taken together, these two processes provided fertile ground for the convergence of the double return and the end-rhyme to a single formal practice: the recapitulatory rotation.

To conclude my discussion, I present further evidence from two types of sources, one textual and the other musical. Both of these amount to little more than circumstantial evidence, but they nevertheless suggest a change of perspective over time that lends support to my view of local practices eventually becoming part of an overarching package.

The first of these is from a recollection by Reichardt of his friendship with J. A. P. Schulz, published in his obituary of Schulz in the *Allgemeine musikalische Zeitung* in 1800:

How often we later laughed over our orthodoxy at that time [the 1770s] with regard to the forms sanctified by the Berlin school. As I played my sonata for Schulz up to the second section of the last movement, he would say to me: Now only a good modulation into a related key is lacking, plus a felicitous return in the main key to a restatement of the preferred places in the first section, and the sonata is *comme il faut*. (Reichardt 1800–1801, quoted in Newman 1963, 599)

William S. Newman (1963, 31) brought this quote as evidence that “as early as 1770 Reichardt and Schulz seemed to have been thoroughly conscious of the recapitulation as being ‘*comme il faut*’ in sonata movements of the Berlin School.” Yet Reichardt’s point was not to tell us how self-conscious he was about using a development and recapitulation in 1770 (in a broader context, it later becomes clear that the point of the anecdote is not how sonata form is constructed but his delight at the use of the French phrase *comme il faut* by a self-proclaimed provincial such as Schulz). On the contrary: he is telling us that, looking back at his works of the early 1770s, he and Schulz *later* became aware of a certain “orthodoxy” in their early works of which they had not been fully aware previously. Reichardt may not have been entirely conscious of the recapitulatory potential of his practice when writing the piece. His description of the “sanctified forms” is, actually, remarkably lax. He does not even explicitly require a double return or an end-rhyme, describing instead two harmonic practices (modulation and return to tonic) and a single subsequent thematic one (the repetition of selected motives), all within a single trajectory covering the entire second half. Reichardt’s works of the early 1770s reflect the flexibility of these requirements: the double return is frequently lacking, the end-rhyme may also not appear, and there is nearly always a good portion of new material after the return to the tonic. But there is always some repetition, and always a well-defined tonic section.³¹ In short, whatever it was that was *comme il faut*, it was certainly not the recapitulation as we know it.

³¹ The works examined here were the first and last movement of the string quartet, string trio, keyboard sonata, and sonata for keyboard and violin from the *Vermischte Musicalien* of 1773. Overall, Reichardt tends to prefer what Hepokoski and Darcy identify as type 2 sonata form.

With hindsight, however, once the recapitulation was an established practice, Reichardt identified his earlier works as having conformed to a more rigid practice than the one he actually practiced at the time. And just as significant, Newman assumed that he was describing a full recapitulation, whereas he was actually describing something more flexible.

Whether or not one accepts my analysis of Reichardt's statement, it is clear that he himself was aware of a perceptual shift regarding formal practices that occurred over the years. A similar perceptual shift (although situated chronologically a little earlier) can be observed in the two-part forms in the keyboard sonatas of C. P. E. Bach from the 1740s and 1750s. Returning to Figure 1, it is easy to understand why this period is of special interest: the double return had only just become a first-level default in binary forms. Furthermore, the end-rhyme had by then become virtually *sine qua non*, so the two features regularly coincided, but without any standard method to combine the two repetition schemata having fully emerged yet. Even a brief perusal of the sonatas of these two decades reveals how many possible solutions were available, even within the works of a single composer, such as C. P. E. Bach. Here I limit my discussion to comparing the first movements of two of his works: the Sonata in B \flat major, Wq. 62/16, composed in 1757 (Example 2), and the Sonata in C major, Wq. 62/7, composed in 1744 (Example 3). Both movements include double returns (mm. 71–76 in the C-major sonata, and mm. 39–41 in the B \flat one), and both movements have substantial end-rhymes (mm. 97–114 in the C-major, and 43–53 in the B \flat). The end-rhyme in the C-major sonata recycles 45 percent of the expositional material, whereas that in the B \flat major sonata recycles 58 percent. Yet the overall logics of the two “recapitulations” (or, to avoid the teleological trap I have been arguing against, the two “post-double-return spaces”) are extremely different.

The B \flat sonata (Example 2), composed in 1757, shows a clearly rotational logic. Following a three-measure double return commencing at m. 39, Bach does not repeat mm. 4–8 (thus omitting the double statement of the opening theme), inserting instead a single measure, the sole purpose of which is to lead smoothly into the end-rhyme in m. 43. The result is a recapitulation that has only one measure that is not “explained” by the exposition. Bach adopts a similar strategy also in the Sonata in E, Wq. 62/17 H. 117, from the same year, this time inserting no new material whatsoever, skipping directly from the half cadence in m. 4 to a tonic version of m. 8, which begins the extensive end-rhyme. The repeated material in both these examples is the *raison d'être* of the entire recapitulation (for a recapitulation is clearly what we have here), and the music after the double return consists almost entirely of repeated or transposed material. In short, these works express a clear sense of an “ideal recapitulation” as a faithful repetition of the exposition, with necessary modifications to suit the tonal requirements.

In the C-major sonata (Example 3), composed thirteen years earlier, the area following the double return is much less indebted to the thematic

Allegro

The image shows a musical score for C. P. E. Bach's Sonata in B major, Wq. 62/16 I. The score is in 3/4 time and B major. It is divided into Exposition (Exp.) and Recapitulation (Recap.) sections. The Exposition (mm. 1-38) features a treble clef with a melodic line and a bass clef with a harmonic accompaniment. The Recapitulation (mm. 39-41) repeats the exposition's material. A second system (mm. 42-45) shows the continuation of the recapitulation, with a treble clef and a bass clef. The score includes dynamic markings such as *p* (piano) and *f* (forte), and articulation marks like accents and slurs. A double bar line with repeat dots is used to indicate the end of the recapitulation section.

Example 2. C. P. E. Bach, Sonata in B[♭] major, Wq. 62/16 I: exposition versus recapitulation

sequence of the exposition. Bach follows the double return with an extensive section based loosely on mm. 18–21 and 23–26 and modeled along the lines of Joseph Riepel’s *Monte* (Gjerdingen 2007, 89–106), traversing the key sequence F, G, C twice: first dedicating four measures to each key center (mm. 77–88) and then summarizing the cycle in three measures (mm. 88–90). Next comes a “false recapitulation” (if it can be called that, coming as it does after the recapitulation) in C minor, and only then does the end-rhyme begin. The space between the double return and the end-rhyme consists of twenty measures, consisting of nearly half the “recapitulation,” that are not justified by any rotational logic and instead have a character more of a development section than of a recapitulation. Research has frequently commented on what Rosen (1988, 289–90) termed the “secondary development,” that is, the use of developmental procedures within the recapitulatory space (normally, but not always, in the transition), which, although not the most common option, was certainly a high-level default in the late eighteenth century (see Wolf 1966, 71; Marston 1992, 794; Hepokoski and Darcy 2006, 235–37; Hoyt 1996). It is frequently viewed as a characteristic of the high classical style,

a kind of evolution from earlier, more formulaic uses of sonata form.³² With C. P. E. Bach's practice in the 1740s, however, recomposing post-double-return space using developmental techniques is quite standard. Typically, as in the example just discussed, this includes some form of Riepel's *Monte*: a quick modulation to the subdominant via the introduction of the flattened seventh (as the dominant seventh or the subdominant of IV), a sequential repeat of the same, transposed upward a major second to modulate to V, and then reinterpretation of V as the dominant, before proceeding to the end-rhyme. The extent of new or loosely adapted expositional material in these examples differs, ranging from a mere handful of measures (Wq. 62/5 I, and most of the Württemberg sonatas) to wholesale replacement, as in the last movement in the Württemberg set, in which twenty-nine measures of new material in the recapitulation replace twenty-one expositional measures in the exposition. The latter example renders a "recapitulation" in which 57 percent of the material is not recapitulated in the sense of strict transposition of expositional material. From both a strictly rotational and a tonal viewpoint, these "free" areas are what Hepokoski and Darcy (2006, 236) would call a "superfluous recomposition"—Bach had already achieved the tonal center in which the end-rhyme will appear, and thus minor adaptations or shortening of the modulatory material in the exposition would have been sufficient. Hepokoski and Darcy argue that both the presence of development-like areas after the double return and the "flat-side" modulations to the subdominant from the 1780s and onward are best understood in dialogical terms: when the "recapitulatory transition" "is given an intense, expanded treatment. . . . The hermeneutic obligation is to explain why" (237), indicating that this is to be viewed as a deformation of more conservative reworkings of transition material. Likewise, their "historically aware hypothesis" explaining the deflection to the subdominant after the double return is that it represents the nonfunctional use of a practice that had originally evolved "over the decades" to serve "manifestly functional purposes" (236). Both explanations may be plausible for post-1780 works, but they cannot be relevant for works composed in the 1740s when no such evolution could be drawn upon. In other words, it is not plausible at a date as early as 1744 to interpret the space beginning at the double return and running to the end of the movement as in dialogue with recapitulatory and developmental procedures, because these latter procedures were not yet well enough established.³³ It seems more reasonable to accept that Bach felt free to make use of the double return and the end-rhyme without any sense of obligation to a faithful reca-

32 For instance, Ethan Haimo (1988, 336) writes that in the 1780s and 1790s Haydn "totally transformed the structure of the recapitulation," whereas Haydn's earlier recapitulations were a "tonally modified repetition of the exposition."

33 Further complication arises from the similarity of these cases to "standard" developments that follow a medial return (Neuwirth 2010). These cases differ from those by C. P. E. Bach only in the existence or nonexistence of a development preceding the double return.

Allegro assai

The musical score is presented in two systems of staves. The first system (measures 1-6) shows the Exposition (Exp.) and Recapitulation (Recap.) sections. A box labeled "Double Return" is placed above the recapitulation staff. The second system (measures 7-12) shows the Exposition (Exp.) and Recapitulation (Recap.) sections. A box labeled "Non-recapitulatory zone" is placed above the recapitulation staff. The third system (measures 13-18) shows the Exposition (Exp.) and Recapitulation (Recap.) sections. A box labeled "Brief -" is placed above the recapitulation staff. Dynamics include piano (*p*) and forte (*f*).

Example 3. C. P. E. Bach, Keyboard Sonata in C major, Wq. 62/7 I: exposition versus recapitulation

pitulatory rotation, but rather with a closer kinship to, perhaps, the kind of thematic reworking one would expect to find before the end-rhyme in baroque binary forms.³⁴ It was only later, as we begin to see in the sonatas of 1757, that he settled upon the solution of modeling the entire space from the

³⁴ The resultant form, with a double return and an end-rhyme separated by new material, has a parallel in ritornello forms, for instance, the first movement of Vivaldi's A-minor Concerto, op. 3/6 (RV 365).

19

Exp.

- thematic parallel

89

Recap.

p

23

Exp.

End Rhyme

Etc.

93

Recap.

Example 3 (continued). C. P. E. Bach, Keyboard Sonata in C major, Wq. 62/7 I: exposition versus recapitulation

double return until the end of the movement as a modified repetition of expositional material.³⁵

These examples lend support to the idea that the conception of the recapitulation as an entirely rotational space was not fully formed (or, at least, was not the only “obvious option”) in Bach’s sonatas of the 1740s, becoming increasingly so toward the 1760s. And although my conjecture is only supported, but not proven, by the developments in Bach’s formal practice, its bottom-up premises provide a sounder and more historically sensitive basis for explaining Bach’s post-double-return procedures than one based on dialogue with a not-yet-formed sonata form.

Conclusion

As shown by the data presented in this article, to understand the rise of the sonata-form recapitulation we must first look separately at its two generating

35 Even in 1757 a strictly rotational recapitulation was not the only solution (even if it was the most frequently used). The sonatas Wq. 62/18 and 62/19 both involve substantially recomposed post-double-return spaces. The former is particularly interesting in that the recapitulation of the

first four-measure phrase is interrupted midway by four measures of developmental procedures, before resuming where it had left off. Furthermore, the end-rhymes are not with the exposition but with the development.

components: the double return and the end-rhyme. Both of these enjoyed an increase in popularity and in length that commenced long before anyone wrote anything that might have resembled sonata form, and only in the mid-eighteenth century did they converge to produce enduring rotational options that enabled the recapitulation to become one of the most prominent features of the classical style. In some sense, their presence in works from the early eighteenth century may be seen as a step on the way to sonata form, but this does not mean that such works should be judged according to the standards of sonata form or evaluated as works in dialogue with it. Instead, by discarding the teleological yardstick of sonata form and the top-down modes of thought inherited from it, we may all the more appreciate the wealth of compositional opportunities available to composers in the first half of the eighteenth century, whose solutions provide provocative suggestions of alternative routes along which classical form might have developed had sonata form not become as dominant as it did. From a methodological point of view, I hope this article has demonstrated the strength of a quantitative approach to form and the ability of such an approach to answer qualitative questions that have the potential to change our understanding of sonata form.

Works Cited

- Allsop, Peter. 1999. *Arcangelo Corelli: New Orpheus of Our Times*. Oxford: Oxford University Press.
- Bonds, Mark Evan. 1988. "Haydn's False Recapitulations and the Perception of Sonata Form in the Eighteenth Century." Ph.D. diss., Harvard University.
- Byros, Vasili. 2009. "Towards an 'Archaeology' of Hearing: Schemata and Eighteenth-Century Consciousness." *Musica Humana* 1/2: 235–306.
- . 2015. "'Haupttrühepunkte des Geistes': Punctuation Schemas and the Late-Eighteenth-Century Sonata." In *What Is a Cadence? Theoretical and Analytical Perspectives on Cadences in the Classical Repertoire*, ed. Markus Neuwirth and Pieter Bergé, 215–51. Leuven: Leuven University Press.
- Caplin, William E. 1998. *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven*. New York: Oxford University Press.
- Caplin, William E., James Hepokoski, and James Webster. 2010. *Musical Form, Forms and Formenlehre: Three Methodological Reflections*. Edited by Pieter Bergé. Leuven: Leuven University Press.
- Churgin, Bathia. 1968. "Francesco Galeazzi's Description (1796) of Sonata Form." *Journal of the American Musicological Society* 21/2: 181–99.
- , ed. 1984. *Sammartini, Giovanni Battista—Ten Symphonies*. New York: Garland.
- Denny, Thomas A. 1988. "Articulation, Elision, and Ambiguity in Schubert's Mature Sonata Forms: The op. 99 Trio Finale in Its Context." *Journal of Musicology* 6/3: 340–66.
- Drabkin, William. 1991. "Beethoven's Understanding of 'Sonata Form': The Evidence of the Sketchbooks." In *Beethoven's Compositional Process*, ed. William Kinderman, 14–19. Lincoln: University of Nebraska Press.
- Gjerdingen, Robert O. 1988. *A Classic Turn of Phrase: Music and the Psychology of Convention*. Philadelphia: University of Pennsylvania Press.
- . 2007. *Music in the Galant Style*. New York: Oxford University Press.

- Haimo, Ethan. 1988. "Haydn's 'Altered Reprise.'" *Journal of Music Theory* 32/2: 335–51.
- Hepokoski, James, and Warren Darcy. 2006. *Elements of Sonata Theory: Norms, Types, and Deformations in the Late-Eighteenth-Century Sonata*. New York: Oxford University Press.
- Hoyt, Peter A. 1996. "The Concept of *Développement* in the Early Nineteenth Century." In *Music Theory in the Age of Romanticism*, ed. Ian Bent, 141–62. Cambridge: Cambridge University Press.
- . 1999. "The 'False Recapitulations' and the Conventions of Sonata Form." Ph.D. diss., University of Pennsylvania.
- Kamien, Roger. 1966. "Style Change in the Mid-Eighteenth-Century Keyboard Sonata." *Journal of the American Musicological Society* 19/1: 37–58.
- Koch, Heinrich Christoph. 1793. *Versuch einer Anleitung zur Composition*. Vol. 3. Leipzig: Adam Friedrich Böhme.
- . 1983. *Introductory Essay on Composition: The Mechanical Rules of Melody, Sections 3 and 4*. Translated by Nancy Kovaleff Baker. New Haven, CT: Yale University Press.
- LaRue, Jan. 1992. *Guidelines for Style Analysis*. Warren, MI: Harmonie Park.
- Lester, Joel. 2001. "Heightening Levels of Activity and JS Bach's Parallel-Section Constructions." *Journal of the American Musicological Society* 54/1: 49–96.
- Mandel-Yehuda, Sarah. 1999. "The Symphonies of Antonio Brioschi: Aspects of Sonata Form." *Min-Ad* 1: 1–8.
- Marston, Nicholas. 1992. "The Recapitulation Transition in Mozart's Music." *Mozart-Jahrbuch 1991*: 793–809.
- Neuwirth, Markus. 2010. "Does a 'Monothematic' Expositional Design Have Tautological Implications for the Recapitulation? An Alternative Approach to 'Altered Recapitulations' in Haydn." *Studia Musicologica* 51/3–4: 369–85.
- . 2013. "Surprise without a Cause? 'False Recapitulations' in the Classical Repertoire and the Modern Paradigm of Sonata Form." *Zeitschrift der Gesellschaft für Musiktheorie* 10/2: 259–91.
- Newman, William S. 1963. *The Sonata in the Classic Era*. Chapel Hill: University of North Carolina Press.
- Rabinovitch, Gilad. 2015. "Tracing Galant Threads: Gjerdingen's Schemata and the Evolution of Musical Form, 1730–1780." Ph.D. diss., Eastman School of Music.
- Ratner, Leonard G. 1956. "Eighteenth-Century Theories of Musical Period Structure." *Musical Quarterly* 42/4: 439–54.
- . 1980. *Classic Music: Expression, Form, and Style*. New York: Schirmer.
- Riepel, Joseph. 1754. *Anfangsgründe zur musicalischen Setzkunst I: De Rhythmpoeia, Oder von der Tactordnung*. Vol. 1. Regensburg: Montag.
- Rom, Uri Binyamin. 2011. "Tonartbezogenes Denken in Mozarts Werken unter besonderer Berücksichtigung des Instrumentalwerks." Ph.D. diss., Technischen Universität Berlin.
- Rosen, Charles. 1988. *Sonata Forms*. New York: Norton.
- . 1997. *The Classical Style: Haydn, Mozart, Beethoven*. Enl. ed. New York: Norton.
- Sisman, Elaine R. 1982. "Small and Expanded Forms: Koch's Model and Haydn's Music." *Musical Quarterly* 68/4: 444–75.
- Strunk, Oliver W. 1974. *Essays on Music in the Western World*. New York: Norton.
- Sutcliffe, W. Dean. 2003. *The Keyboard Sonatas of Domenico Scarlatti and Eighteenth-Century Musical Style*. Cambridge: Cambridge University Press.
- Talbot, Michael. 1996. "Stylistic Evolution in Corelli's Music." *Studii Corelliani* 5: 143–58.
- Tymoczko, Dmitri. 2011. *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice*. New York: Oxford University Press.
- Webster, James. 1982. "Binary Variants of Sonata Form in Early Haydn Instrumental Music." In *Proceedings of the International Joseph Haydn Congress*, ed. Eva Badura-Skoda, 127–34. Munich: Henle.

- . 2001. "Sonata Form." In *The New Grove Dictionary of Music and Musicians*, 687–98. www.oxfordmusiconline.com/subscriber/article/grove/music/26197 (accessed October 22, 2013).
- Wolf, Eugene K. 1966. "The Recapitulations in Haydn's London Symphonies." *Musical Quarterly* 52/1: 71–89.
- . 1981. *The Symphonies of Johann Stamitz: A Study in the Formation of the Classic Style, with a Thematic Catalogue of the Symphonies and Orchestral Trios*. Vol. 2. Utrecht: Bohn, Scheltema en Holkema.
- Yule, G. Udney. 1912. "On the Methods of Measuring Association between Two Attributes." *Journal of the Royal Statistical Society* 75/6: 579–652.

Yoel Greenberg is lecturer in music at Bar-Ilan University and violist with the Carmel Quartet. His research interests concern the evolution of sonata form and interactions of music, art, and literature in the early twentieth century. ♣