ZGMTH

Surprise Without a Cause?

'False Recapitulations' in the Classical Repertoire and the Modern Paradigm of Sonata Form[1]

Markus Neuwirth

ABSTRACT: False recapitulations are often cited as a hallmark of Joseph Haydn's sonata-form style, exemplifying perhaps better than any other technique the composer's witty and subversive engagement with formal conventions. However, closer scrutiny reveals that the concept of false recapitulation is based on a number of different, partially incompatible cognitive, intentional, theoretical, and historical criteria. In an attempt to reconstruct the horizon of expectations of historical listeners, I shall essentially draw on two sources: the compositional practice of the time as reflected in a preliminary repertoire study and contemporaneous theoretical writings. In a nutshell, I shall argue that the analytical practice of framing a double return in the development section in terms of a play with listener expectations is based on the anachronistic assumptions of what I call the "modern paradigm of sonata form". Placing expectations at the center of analysis and scrutinizing its complex preconditions allows us to arrive at a refined understanding of Haydn's (and others') usage of supposedly false recapitulations.

Falsche Reprisen werden gerne als ein wesentliches Kennzeichen von Joseph Haydns Sonatenformen angeführt; sie exemplifizieren vermutlich besser als jede andere Formstrategie die „witzige" und subversive Auseinandersetzung mit formalen Konventionen. Allerdings zeigt eine gründliche Prüfung, dass das Konzept der falschen Reprise auf einer Reihe unterschiedlichster, teils inkompatibler Kriterien kognitiver, intentionaler, theoretischer und historischer Art beruht. Bei dem Versuch einer Rekonstruktion des Erwartungshorizontes historischer Hörer stütze ich mich im Wesentlichen auf zwei Quellen: die kompositorische Praxis der Zeit, wie sie sich in einer vorläufigen Repertoirestudie widerspiegelt, sowie historische musiktheoretische Schriften. Im Kern wird argumentiert, dass die analytische Praxis, die Rückkehr des Hauptthemas in der Grundtonart im Verlauf der Durchführung als ein Spiel mit Hörerwartungen aufzufassen, auf anachronistischen Annahmen beruht, die sich aus dem speisen, was ich als „modernes Paradigma der Sonatenform" bezeichnen möchte. Der analytischen Fokus auf Hörerwartungen und die genaue Prüfung von deren komplexen Vorbedingungen ermöglicht es, ein wesentlich verfeinertes Verständnis von vermeintlich falschen Reprisen zu erlangen, wie sie von Haydn und seinen Zeitgenossen gebraucht wurden.

Introductory Remarks

Although studied extensively by music psychologists, 'expectation' (or 'expectancy'[2]) is no doubt a neglected category in current music-analytical writings. This diagnosis might come as a surprise to most readers since numerous music-analytical studies, whether of tonal or posttonal music, refer to listening expectations in one way or the other. However, the crucial problem is that this is often done implicitly, without any deeper reflection on the underlying cognitive foundations and preconditions. A notable exception in this regard is the implication-realization theory originally devised by Leonard B. Meyer and further developed by Eugene Narmour[3]—a theory, however, that did not gain significant currency in the analytical practice. Furthermore, Meyer's theory focuses solely on melodic processes as exemplified by the so-called changing-note schema (e. g., 1–7–4–3 in the soprano) and the gap-fill pattern.[4] Other approaches, which are more psychological (or experimental) in nature, are equally restrictive in scope, paying attention to harmonic or metrical (or rhythmical) expectations exclusively.[5] Overall, one notices a striking parametric limitation imposed on the concept of musical expectancy;[6] a holistic concept is conspicuously lacking.[7]

That expectation is one of the most central analytical categories is testified not only by the frequent invocation of listener expectations in analytical studies of individual compositions,[8] but also by the fact that hypothesized expectations are strikingly manifest in the analytical terminology. A familiar example of the latter is the concept of 'deceptive cadence,' which implicitly expresses the expectation of a normative cadential progression, one ending on the tonic.[9] Also in the realm of musical form—a level that the present article seeks to address—analysts often more or less overtly refer to listening expectations when pointing out the non-normative status of a certain formal strategy. Relevant examples include movements beginning on harmonies other than the tonic, 'false' transitions (which raise the expectation of, but ultimately fail to lead to, the subordinate theme and key),[10] subversive strategies in dealing with the medial caesura in sonata-form

---

2 Eerola 2003.

3 Meyer 1956, 1967, and 1973; Narmour 1977, 1990, 1992, and 1999. Further studies of melodic expectations include Larson 2002 and Margulis 2005.

4 The latter can more easily be explained as the result of a tessitura effect, see Hippel / Huron 2000.

5 For studies of harmonic expectations, see, for instance, Bharucha 1987; for rhythmic expectations, see Jones 1987 and Schmuckler 1989.

6 Characterizing a general trend in music psychology, Clarke (1989, 4) coined the term "parametric separatism."

7 A notable exception is Huron 2006.

8 See, for instance, Danuser 1986, Bandur 2002, and Hinrichsen 2004.

9 Cf. Neuwirth (in preparation).

10 Hepokoski / Darcy 2006, 80. Such a 'false transition' arises when, subsequent to a complete main theme, a passage enters that dynamically and texturally (as well as with regard to a loose-knit phrase structure) evinces unequivocal characteristics of a transition, but eventually fails to prepare the entrance of the second theme in the new key and instead returns to the main theme in the tonic key. (This typically happens after a half cadence in the tonic key.) Accordingly, the caesura appearing after the half cadence is dubbed "false medial caesura" or "false I:HC MC" (Hepokoski / Darcy 1997, 140).

expositions,[11] and 'false retransitions' (failing to prepare for the entrance of the recapitulation proper), as well as 'false,' 'off-tonic,' and 'veiled' recapitulations.[12] In all these cases, the qualification 'false' is usually chosen to indicate that we are confronting an intentional game upon listening expectations on the part of the composer—a game based on a presupposed formal norm as the background premise for expectations to arise.

In this article, I will restrict myself to discussing the so-called 'false recapitulation.' False recapitulations are often cited as a hallmark of Joseph Haydn's sonata-form style, exemplifying perhaps better than any other technique the composer's witty and subversive engagement with formal conventions. Nevertheless, as this article aims to show, closer scrutiny reveals that the concept of false recapitulation is based on a number of different, partially incompatible cognitive, intentional, theoretical, and historical criteria.[13] In an attempt to reconstruct the "horizon of expectations" (Gadamer) of historical listeners, I shall essentially draw on two sources: the compositional practice of the time as reflected in a preliminary repertoire study (see 2.) and contemporaneous theoretical writings (see 3.). In a nutshell, I shall argue that the analytical practice of framing a double return in the development section in terms of an intentional play with listener expectations is based on the anachronistic assumptions of what I call the "modern paradigm of sonata form." Placing expectations at the center of analysis and scrutinizing its complex preconditions allows us to arrive at a refined understanding of Haydn's (and others') usage of supposedly false recapitulations.

1. The Modern Paradigm of Sonata Form

Expectations never emerge in a vacuum. Rather, they are invariably guided by some sort of (implicit) theory. When speaking of expectations as implied in the notion of false recapitulation, it is therefore mandatory for any analytical endeavor to clarify the underlying theoretical premises. The current theoretical paradigm, widely accepted no later than the second half of the twentieth century in both English- and German-language countries, emphasizes the form-defining role of tonal processes and cadential markers.[14] In particular, this widely shared paradigm is essentially based on the introduction and resolution of a large-scale tonal tension. By moving away from the home key and subsequently establishing and confirming the subordinate key (the key of V), the exposition creates a 'large-scale dissonance'[15] that calls for resolution in the movement's second half.[16]

11 For a critical review of the idea that continuous expositions lacking a medial caesura should be understood as a play on the convention of the two-part exposition, see Neuwirth 2011.

12 See Neuwirth (in preparation).

13 For critical reviews of the concept of false recapitulation, see also Hoyt 1999, Kim 2004, Burstein 2009 and 2011a, and Korvstedt 2013.

14 E.g., Tovey 1944, Ratner 1949, Larsen 1963, Rosen 1988, Webster 2001, and Hepokoski/Darcy 2006.

15 See Rosen 1988, 229.

16 Rosen 1988, 244. There are some obvious conceptual limitations to this idea: transferring the concept of dissonance from the level of chords to the level of keys and thus treating different hierarchical levels of structure in a uniform manner is not unproblematic, since a fifth (or a third, in minor-

The resulting tonal polarity can optionally be emphasized by a contrast at the thematic level.[17]

The (expendable) function of the central section, the development, in this paradigm is seen to prolong V and thus to intensify the state of tension by effectively delaying the entrance of the double return (i. e., the simultaneous occurrence of the primary theme and the home key) and the subsequent tonal resolution.[18] The development is therefore described as "a (gigantic) transition from the end of the exposition to the beginning of the recapitulation."[19] In order not to weaken the extraordinary psychological effect conveyed by the tonic return at the moment of recapitulation, the tonic should be consistently avoided beforehand.[20] The resolution of the large-scale dissonance and thus the 'structural' and 'psychological climax'[21] of the entire movement is persevered for the recapitulation launched by the double return. The double return may be prepared by an emphatic dominant sonority at the end of the developmental retransition.[22]

Yet if the retransition is said to prepare resolution, the crucial question is at which point in the movement the resolution occurs. The majority of theorists would agree that, in Rosen's words, "the *essential* resolution is that of the second group which has never been played in the tonic and must be so played before the piece can be considered over, the matter closed."[23] By transposing the 'secondary theme-closing theme' block down a fifth or up a fourth (or mixing both methods), the composer succeeds in fulfilling the requirements of the 'sonata principle.'[24] Note that also from a Schenkerian point of view, the moment of double return is not understood as a moment of resolution but rather as

---

mode movements) in traditional counterpoint is normally not considered a dissonant interval that would demand resolution. Because of the different properties of these levels of musical structure, one runs the risk of committing what Leonard Meyer has aptly called the "fallacy of hierarchical uniformity" (Meyer 1967, 96) when speaking of a large-scale dissonance. To use this expression is legitimate only insofar as the analogy between dissonant chords and keys seeks merely to emphasize the fact that the relationship between the tonic and the subordinate key is one that needs to be problematized, balanced, or corrected.

17 However, this contrast does not acquire the form-functional relevance it had in the by-now obsolete thematic (dualistic) nineteenth-century model.

18 See Rosen 1988, 262f.

19 Webster 2001, 688.

20 See, for instance, Stöhr 1927, 370 and Leichtentritt 1927, 141 (Leichtentritt 1951, 134). This view has been anticipated by Czerny 1848, 36: "As to the modulations in the development of the second part, the composer has a free choice of all keys. But he must, to a certain extent, avoid the original key of the piece, and that of its dominant, so as not to dwell in them for any length of time, or to employ them for any considerable idea, because they have been sufficiently used in the first part."

21 Cf. Webster 2001, 692. By replacing "psychological" through "structural" in the second edition of the *New Grove Dictionary*, Webster suggests the interchangeability of these two terms and hence the psychological relevance of the structural process.

22 Cf. Rosen 1988, 262f. On the psychological significance of the retransition, see, for instance, Rywosch 1937, 133. The retransition frequently makes use of what Schoenberg called an "upbeat chord" (see Schoenberg 1967, 209).

23 Rosen 1988, 157 (my emphasis).

24 See, for instance, Cone 1968, 76f., Haimo 1995, 3, Webster 2001, 688, and Hepokoski 2002.

articulating a "middleground rebeginning."[25] The resolution of the background dominant takes place later in the form, namely at the end of the retransposed and cadentially confirmed secondary-theme group.[26] Here, the *Urlinie* reaches its goal, the first scale degree, which has previously been delayed.

So if the double return itself cannot be considered a "point of repose,"[27] why then should we ascribe any structural significance to this moment? Two aspects are worth mentioning here: First, this moment may be considered significant because it creates a parallel to the beginning of the movement, or, in Ratner's words, "a *rhyme* of the melodic material of part I"[28]: the double return is the initial event of a (potentially complete) rotational cycle through the thematic material of the exposition.[29] Second, harking back to a strategy inspired by semiotics, the moment of double return can also be understood as a prospective musical sign, one that *signals* a moment of resolution yet to come but which it does not bring about itself. Specifically, it functions as a sign pointing ahead to the upcoming moment of resolution at the structural cadence, the I:PAC occurring at the end of the transposed secondary-theme group. That the double return can act as a *pars-pro-toto* for the entire formal section called "recapitulation" is possible because the double return has been associated with tonal resolution so frequently in the history of sonata form.[30]

2. The traditional view of the 'false recapitulation' and its underpinnings

Given the premises of the modern paradigm of sonata form, the simultaneous return of primary theme and home key in the development section—a phenomenon Peter Hoyt dubbed a "medial double return"[31]—seems to be a deeply problematic event, explica-

25 Smith 1994, 79. The direct resolution of the retransitional dominant to the adjacent tonic at the start of the reprise is merely a local event that takes place at the foreground level (see Burstein 2005, 306–308). It is only from the perspective of a higher level of abstraction—the middleground—that the retransitional V at the end of the development articulates a moment of interruption of the descending *Urlinie* on the second scale degree. With the entrance of the recapitulation, the *Urlinie* reverts to the *Kopfton* (2–3 or 2–5).

26 Cf. Hepokoski/Darcy 2006, 147.

27 Bonds 1988, 330.

28 Ratner 1980, 229.

29 On the notion of rotation, see Hepokoski/Darcy 2006, 611–614. The primary theme's basic idea might potentially be sufficient to generate the expectation of a full recapitulation, projecting into the future the events we encountered as the exposition unfurled, an option that Hepokoski und Darcy refer to as "synecdochic strategy" (ibid., 233).

30 Peter Hoyt offers a very concise account of the underlying (tacit) semiotic premise of many sonata-form approaches that speak of the *onset* of the recapitulation as a moment of resolution; see Hoyt 1999, 16.

31 Ibid., 43. Hoyt's term is meant to encompass not only those double returns that are traditionally considered to be potential false recapitulations, but also those that should be viewed as having a ritornello-like meaning (as associated with the convention of what Bonds refers to as "precursory return" [Bonds 1988, 220–224]). Some alternative terms can be found in the literature; however, they are not necessarily neutral in their implications. For Strunk, the term "premature reprise" was intended to denote true recapitulations (as opposed to a variant of the false recapitulation (1932, 236f.). Rosen offers a critical view of this term: "It can only be considered premature with respect to

ble only in terms of the intentional play upon, and deception of, listening expectations, in short: a false recapitulation.[32] The qualification 'false' indicates that this occurrence is not definitive: The double return is not followed by the remaining expositional units that would be needed for a complete recapitulatory rotation (i. e., there is no long-range *thematic parallelism*), nor does the subsequent passage stay in the tonic key but instead modulates to new keys, most typically the submediant (i. e., there is no *tonal unity* at this point[33]). By quickly moving away from the tonic and resuming genuinely developmental textures and techniques, composers unequivocally signal to their audience that the initial interpretation of the P-restatement as the moment of recapitulation was in fact premature.[34]

Apart from conceiving of the false recapitulation as a play upon listening expectations, analysts, especially those adhering to Schenkerian theory, have developed a second immunization strategy: they propose the idea that a large number of double returns located in the development section can best be understood as 'apparent tonics.' Though *materialiter* giving the impression of a tonic, when viewed from a higher (or more abstract) structural level, the tonic arrival co-occurring with the thematic return in these cases is merely a transient phenomenon, one that is embedded in a larger harmonic progression.[35] Movements featuring true returns of the tonic *Stufe* early in the development must, in Oster's words, be considered "borderline cases of sonata form."[36] This understanding is consistent with the basic axioms of the Schenkerian *Schichten* model: The tonal "consonance" produced by the medial double return seems to represent a challenge to the modern paradigm of sonata form, which, as we have seen, is based on the idea of a long-range tonal dissonance and its subsequent resolution. If the tonic return in the development acted as an emphatic *Stufe* (as opposed to a local chord) within the overall voice-leading structure, the teleological (or dramatic) design of the sonata-form movement would potentially be jeopardized. In an attempt to resolve this problem, Schenkerian theorists have argued that tonic returns in the development section act as

the model that was to become canonic much later" (1988, 155). Despite the fact that Rosen recognizes the anachronistic implications inherent to this term, he is reluctant to propose any alternative, "as there is no point in multiplying terminology" (ibid., 276). Webster and Haimo use the term "immediate reprise" instead (Webster 1986, 128; Haimo 1995, 105).

32 The issue of whether an off-tonic thematic return can function as a false recapitulation is discussed at some later point in this article.

33 If the music that follows the double return stays essentially in the tonic key (except for brief, structurally insignificant digressions), such that S is transposed into the tonic and, concomitantly, structural closure is achieved, the initial hypothesis that the double return articulated the onset of the recapitulation is eventually confirmed (Bonds's criterion of the "definitive re-establishment of the tonic;" [1988, 207]).

34 In addition to the deceptive effect achieved by the false recapitulation, there might be a genuinely structural element to it as well, namely "lengthening the development" and hence clarifying the ternary structure of the sonata form, as Robbins Landon has observed (1955, 303).

35 As early as in his *Harmonielehre* from 1906, Schenker cautions against confusing scale-step and triad: "not every triad must be considered as a scale-step […] The scale-step is a higher and more abstract concept" (Schenker 1954, 138f.).

36 Schenker 1979, 140.

'apparent' or 'pseudo' tonics that differ from tonic *Stufen* in that they fulfill an entirely different function with regard to the large-scale tonal structure.[37] Unlike *Stufen*, which are situated in the deep middle-ground of a movement's structure, apparent tonics are transient (subordinate) and relatively brief events located in the structural foreground.

The fact that the interpretation of a given tonic return in the development as a *Stufe* also depends on the duration of that event becomes clear from Burstein's analysis of Haydn's Symphony No. 55,i. Here, the double medial return is too extensive to be understood as a purely transient event.[38] So if one defines sonata form primarily in terms of a *single* moment of interruption, as Schenker did, then one would indeed be forced to deny the label "sonata form" to this movement. Such a conclusion would be entirely consistent with the logic of the hierarchic approach that lies at the heart of Schenkerian theory: When two events compete with one another with regard to their structural importance, the hierarchical approach forces a clear-cut decision. The crucial premise inherent to Schenkerian theory is that there can be only one true tonic *Stufe* at the beginning of the recapitulation to which everything else is subordinate. Abandoning the criterion of singularity (by assuming that this event might occur more than once) could potentially endanger the *Schichten* model upon which Schenkerian theory rests.

Apart from this problem, it is important to note that the status of the tonic return within the large-scale structure becomes evident to the listener only in retrospect. At the moment of its sounding, i. e., from the perspective of the foreground level, a double return appearing in the development section may indeed evoke a feeling of surprise. In continuing to use the term 'false recapitulation,' Schenkerians intend to express this temporary and local effect of deceptiveness conveyed by the double return. What they would deny, however, is that the tonic return in the development articulates a moment of *true* consonance *pace* Rosen.[39]

Despite the familiarity of both the traditional account and its Schenkerian variant, to speak of a musical event as a false recapitulation is a much more complex and loaded statement than one might tend to think at first glance, one that involves a number of inter-related cognitive, intentional, theoretical, and historical assumptions.

(1) The concept of false recapitulation implies the notions of expectation and expectancy violation (surprise). Generally speaking, expectations result from an interplay between specific stimulus situations and mentally represented, learned stylistic conventions, as has been argued by Leonard B. Meyer in his seminal study entitled *Emotion and Meaning in Music* (1956).[40] In other words, expectations feature a subjective and an objective component: as to the former, the formation of expectations presupposes the psychological process of knowledge acquisition, a fact that prompts us to consider the cognitive issue of statistical learning of these conventions by the listener (in the sense of Meyer's understanding of style as an "internalized probability system"[41]). As to the lat-

---

37 For a general account of the notion of "apparent tonic," see, for instance, Adrian 1990 and 1991.

38 Burstein 1999, 78 (note 13).

39 See Rosen 1988, 280.

40 Meyer 1956, 72.

41 E. g., Meyer 1967, 47.

ter, analysts invoking the notion of expectation inevitably touch on the issue of stylistic regularities, norms, and conventions; they are forced to be as precise as possible about the stylistic framework within which expectations are formed, while being aware of the difficulties involved in trying to get access to the horizon of expectations of historical listeners.[42]

An additional complication lies in the dynamic nature of expectations: expectations are not static and constant but change over time as a function of shifting contexts and time-windows—an aspect forcing the analyst to consider music as a temporal process unfolding in real-time (as opposed to an atemporal, spatial object).[43] This dynamic cannot only be found within a given piece, but also across a large number of pieces distributed over a certain period of time. The norms upon which listening expectations rest are not given to us *a priori* but develop, crystalize, and disappear over time.

Consider the following quotation. Referring to Haydn's Symphony No. 55,i, Poundie Burstein argues that "the appearance of the so-called 'false recapitulation' of bars 97-102 should not be regarded as an example of humor [Rosen's view; M.N.]. Haydn used this device frequently in his symphonies written around this time. As a result, *listeners sensitive to his style* would not be surprised by the 'false recapitulation,' but would recognize it as a standard feature of Haydn's symphonic sonata form."[44] This argument suggests that at one point in the history of sonata form, listeners must have recognized that the false-recapitulation strategy was used with such regularity that it became a convention of its own and hence in a way expected. From that moment on, the false recapitulation would have been effective only if one disallowed for any learning process on the part of the listener.

But does sensitivity to a given style or convention necessarily preclude the experience of deception or violation of listening expectations? There seems to be the paradox that when speaking of expectations, an event may be at once surprising and expected, depending on the level of description one chooses. This type of problem has already been illuminated, in more abstract terms, by Naomi Cumming. She has pointed out the contradiction between the claim that listeners have acquired stylistic competency and the phenomenon of deceived listening expectations. Meyer's assertion that listening expectations rely on, and are indicative of, the internalization of stylistic norms can only be maintained on the basis of a differentiation of the concept of style: Cumming argues that only the abstract rules (derived from a given corpus) have been mentally internalized, not the individual solutions and strategies serving the realization of these rules.[45]

---

42 The context-sensitivity of expectations is reflected in the following statement by Meyer (1956, 30): "[…] the same physical stimulus may call forth different tendencies in different stylistic contexts or in different situations within one and the same stylistic context. For example, a modal cadential progression will arouse one set of expectations in the musical style of the sixteenth century and quite another in the style of the nineteenth century."

43 Such processual approaches, invoking the notion of retrospective reinterpretation with regularity, are adopted in Federhofer 1981, Lewin 1986, and Dahlhaus 1987.

44 Burstein 1999, 78 (note 13 [my emphasis]).

45 Cf. Cumming 1991, 185. A variant of the problem that Cumming describes is the so-called "Wittgenstein paradox" (Dowling and Harwood's term; 1986). If a listener is completely familiar with a

Another important aspect of the focus on expectations implied in the notion of false recapitulation is that it allows to re-integrate the (historical and modern) listener in the analytical process. When invoking the concept of listener at all, it often remains unclear what type of listener analysts are referring to. Typically, music analysts conceive of the listener as an "ideal" one (sometimes in analogy to the "implicit reader" invoked in the Iserian reader-response or reception theory).[46] According to Erwin Ratz, an "ideal listener" is highly competent and attentive, someone who notices everything, but who does not know what is about to come next. In terms of Bharucha's useful distinction, one might say this type of listener does not possess any 'veridical' (or 'intra-opus') but only (perfect) 'schematic' (or 'extra-opus') knowledge.[47] The mode of listening Ratz describes is the one that listeners experience upon first exposure to a given piece of music.[48] In a similar vein, Adorno, who in his *Introduction to the Sociology of Music* (1976) develops a typology of listeners, defines what he calls the "expert listener" as "the fully conscious listener who tends to miss nothing and at the same time, at each moment, accounts to himself for what he has heard. […] Spontaneously following the course of music, even complicated music, he hears the sequence, hears past, present, and future moments together so that they crystallize into a meaningful context."[49] So for both Ratz and Adorno, this type of listener does not only possess extraordinary musical skills but also a remarkably powerful memory, storing all the information contained in a given piece and being capable of connecting this information to new one. Taking "the standpoint of a listener who knows nothing beforehand, but hears and remembers everything" amounts to what Tovey calls "a true analysis."[50] It is precisely this type of listener that analysts seem to have in mind when invoking the concept of false recapitulation. Especially the extraordinary memory capacities that the ideal listener is assumed to possess surfaces in the following two more specific premises underlying the notion of false recapitulation.

(a) Whether or not a double return is considered to articulate a true or false recapitulatory beginning essentially depends on what follows this event. Until more context is provided to the listener, either interpretation is nothing but a mere hypothesis. That also the formal status of a double return acting as the initial event of a true recapitulation may at times be ambiguous, owing to the continuation of developmental activity (in the sense

given piece, this familiarity is seen to preclude any feeling of surprise. As a result, the piece under consideration would be deprived of its aesthetic qualities. To make a useful analogy, why is it, for instance, that a deceptive cadence in a familiar piece still sounds unexpected? This problem can only be resolved by distinguishing between two different types of expectations, those based on veridical memory and those based on schematic memory. Crucially, these two memory systems are assumed to be modularly separated.

46 La Motte-Haber 1995, Butt 2010, and Hepokoski / Darcy 2006, 605–610.

47 Bharucha 1987. As style is essentially defined by Meyer (1989, 3) as resulting from replicated patterns, arising from the composers' constrained choices among options from a more or less fixed set, 'intra-opus' style refers to such patterns replicated within the same composition, whereas 'extra-opus' (or 'inter-opus') style denotes bundles of musical features reiterated across a given number of compositions.

48 Ratz 1968, 8.

49 Adorno 1976, 4.

50 Tovey 1935, 68.

of a 'second development'), becomes evident from Bonds's description of the concept of 'false-false recapitulation': "After re-introducing the tonic and the main theme, the composer often abandons both and in effect *sustains the impression of development*. At times, the departure from the re-established tonic is so swift that one is forced to question the solidity of the return to the principal key and the beginning of a new section within the sonata-form movement."[51]

Although Bonds acknowledges the difficulty of "defin[ing] precisely which works employ this technique," he cites the finale of Haydn's Symphony No. 54 (1774) as "a good example of a decidedly ambiguous return […]."[52] A characteristic feature of this example is that "[o]nly in retrospect, with the return of S in the tonic (m. 128) does the structural meaning of this passage become clear. What may at first sound like a false recapitulation ultimately emerges as the true one."[53] This points to two further important criteria underlying our notion of recapitulation. First, the *singularity* of the double return: If the double return is not followed by another (non-adjacent) P-restatement (except in the coda) that could possibly challenge the double return's formal status (and considerably diminish its psychological impact), the double return should be understood as the onset of a true (as opposed to false) recapitulation. Second, the retrospective evaluation of the validity of a particular formal hypothesis: Only when the course of events following the (tonic or off-tonic) return of P closely resembles the events previously heard in the exposition, and when there is no subsequent tonic return of P, will the listener be inclined to grant the thematic restatement recapitulatory status. However, in principle, this is only possible in retrospect[54] and further requires the listener to be able to hear, in Adorno's words, "past, present, and future moments together so that they crystallize into a meaningful context."

(b) A second cognitively relevant issue concerns the key(s) in which the primary theme can appear in order to deceive listeners with respect to its formal status. One might argue that P-appearances in keys other than the tonic are not generally considered similar enough to what we would expect at the moment of true recapitulation to be viewed as viable formal rebeginnings. Although this argument may seem self-evident, there is some controversy surrounding the issue of the tonality of false recapitulations.

---

51 Bonds 1988, 322 (my emphasis).

52 Ibid. The following list (see Neuwirth, in preparation) includes further examples of this technique, some of which feature cadences in keys other than the tonic (1), while others do not (2): (1) bIII:PAC in Haydn, Keyboard Trio Hob. XV:21/i; bIII:IAC in Haydn, Symphony No. 54/iv; and in Woelfl, Violin Sonata op. 16 No. 2/i; bIII:HC in Haydn, Keyboard Trio Hob. XV:11/i and in Mozart, K. 485; iii:HC in Haydn's Keyboard Trio Hob. XV:14/iii. (2) C. P. E. Bach, Sonata No. 4 in B-flat major, W49/4; Haydn, Symphony 89/i; Hob. XVI:24/ii; Hob. XVI:29/i; Hob. XV:13/ii; Hob. XV:16/i; Hob. XV:21/iii; Haydn, opp. 9 No. 2/iv, 9 No. 5/iv; 17 No. 1/i, 20 No. 5/i, 64 No. 4/i; Koželuch, Keyboard Sonatas opp. 15 No. 2/i, 26 No. 1/i, and 35 No. 2/i; Mozart, Keyboard Trio K. 542/i.

53 Bonds 1988, 322ff. On the notion of false-false recapitulation, see Larson 2003, 143 (on Haydn's String Quartet op. 2 No. 4,i) and 155 (on Haydn's String Quartet op. 17 No. 4,i). Larson also uses the term "disguised recapitulation" as a synonym for "false false recapitulation" (ibid., 155 and 159, on Haydn's String Quartet op. 20 No. 1,i). Rosen views the "false false-reprise" as "an exceptionally sophisticated irony even for Haydn" (1997, 141).

54 See also Tobel 1935, 168.

Unlike Mark Evan Bonds, who adheres to a narrow definition of the concept, including only tonic-key instances of P,[55] James Webster proposes a broader definition of the term, also classifying main themes in other keys as false recapitulations. Such non-tonic (or 'wrong'-key) false recapitulations, which Leichtentritt refers to as *Scheinreprisen*,[56] are said to occur frequently in the late works of Haydn, e. g., in his B-flat major Symphony No. 102,i (mm. 185 ff.).[57] Following Webster, William Caplin likewise defines a false recapitulation as "the appearance of main-theme material in a tonal region other than tonic of the home key."[58] Since this typically occurs "[n]ear the end of a development or a rondo couplet,"[59] false recapitulations retrospectively reveal themselves as retransitions, starting "with reference to the opening material from the main theme, usually in the development key just confirmed by a prior half cadence."[60] Retransitions that prepare false recapitulations are themselves referred to as 'false,' since they raise expectations that are ultimately frustrated.[61]

Such broader definitions of the term seem to be problematic for two reasons: First, as some authors have argued, composers frequently communicated to the listener the preliminary formal status of such thematic returns not only by relying on tonality alone. In other words, the off-tonic key *by itself* is not sufficient to cast doubt on the recapitulatory function of the thematic return, since this is a component that can only be perceived by listeners with relative pitch and a (near) perfect tonal memory. Taking into account the limited cognitive capacities of their listeners, composers chose to express this through readily comprehensible surface features that are generally considered atypical of genuine recapitulatory beginnings, such as a thin texture, soft dynamics, and reduced orchestral forces.[62]

Second, such broader definitions are also problematic in the light of Rosen's famous statement that "[a] false reprise is not only a false resolution, but a brief moment of consonance in the most dissonant section of the work."[63] Rosen's characterization could be interpreted as implying that false recapitulations must appear in the tonic key in order to be able to produce a moment of "consonance." Yet, presumably to account for the phenomenon of subdominant recapitulations, Rosen also makes allowances for false recapitulations that enter in the key of IV: "[…] if a reprise is not in the tonic (*or the sub-*

---

55 Cf. Bonds 1988, 229.

56 Cf. Leichtentritt 1927, 162.

57 Webster 2001, 693.

58 Caplin 1998, 254.

59 Ibid.

60 Ibid., 159.

61 Sisman defines the term "false retransition" as follows: "Thus, what is critical here [in the opening movement of Mozart's Jupiter Symphony] is that a recapitulation is *expected*, not that the listener is momentary misled when it appears. […] What is false here – what misleads the listener – is really the retransition, since there is palpably no recapitulation. […] We might consider the false retransition and recapitulation-interlude to be 'formal' topics – topics of interior reference – rather than exterior topics, such as those derived from the dance" (1993, 52 f.). See also Spitzer 1996.

62 Levy 1982. Thomson describes this situation as a case of synchronic ambiguity (1983, 10).

63 Rosen 1988, 280.

dominant), it fools only the uneducated […]."[64] In making such a definitive statement, Rosen implicitly criticizes all those theorists who—like Webster or Caplin—claim that main themes in keys other than the tonic or the subdominant can potentially mislead listeners with respect to their formal status.

Yet Rosen's argument presupposes a listener's abilities to differentiate between keys and to recognize large-scale tonal relationships. It is of course speculative, if not downright impossible, to answer the question of whether historical listeners possessed these abilities. Present-day listeners, however, seem to lack them, as a number of empirical studies suggest.

In an oft-cited study, Nicholas Cook concludes from his experimental results "that the tonal unity of a sonata is of a conceptual rather than perceptual nature, in contrast to the directly perceptible unit of a single phrase. […] the theories of Schenker, Meyer, and Lerdahl and Jackendoff are better seen as a means of understanding the practice of tonal composers than as a means of predicting the effects of their compositions upon listeners."[65] Although Cook's study has rightly been criticized for a number of serious methodological flaws[66], later studies essentially replicate the general finding that listeners are largely insensitive to global harmonic structures.[67] Some of these studies even acknowledge the possibility that tonal memory may draw on indirect perception (or inferences) based on familiar stylistic features that are located at the musical surface and hence are readily recognizable.[68] As Cook himself concedes, "if large-scale tonal relations are not themselves audible, that does not necessarily mean that they are of no musical significance: it may just be that their influence on what is heard is an indirect one."[69] And what is heard on the surface includes "[t]he disposition of textures and thematic materials, the patterning of loud and soft passages and of high and low tessitura, the pacing of tension and relaxation."[70] For Cook, "all these aspects of a sonata are organized around the tonal plan and serve to project its structural closure in a directly perceptible manner."[71] If we agree that eighteenth-century listeners shared with modern listeners the insensitivity to long-range tonal relationships, composers of the time could rely only on secondary parameters to communicate form on a local level. In other words,

---

64 Ibid., 282 (emphasis in original). Examples of subdominant recapitulations can be found in the works of Gassmann, Dittersdorf, Haydn, Koželuch, Clementi, Mozart, Pleyel, und Schubert (for an extensive list, see Neuwirth, in preparation).

65 Cook 1987, 204.

66 See Gjerdingen 1999. The most important criticism concerns the problem that Cook actually measures aesthetic preference (including coherence and pleasure) rather than the perception of tonal unity, suggesting that the former is a consequence of the latter.

67 E. g., Tillmann/Bigand/Madurell 1998.

68 E. g., Marvin/Brinkman 1999. Factors determining listeners' abilities to recognize long-range tonal relationships include (1) expertise (experts/musicians vs. novices/non-musicians); (2) relative/absolute pitch (tonal memory); (3) duration (temporal distance); (4) key distance; and (5) familiarity with a given stimulus (piece) as a result of repeated exposure.

69 Cook 1987, 204.

70 Ibid. See also Burstein 2011b, 118.

71 Cook 1987, 204.

they could enable a mode of listening that Maus calls "hearing of large-scale form in the moment"[72] by using surface features such as "texture as a sign" indicating large-scale formal functions.[73] This assumption is plausible also in view of the hypothetical nature of any thematic return: Unless confirmed by the remaining events of the recapitulatory rotation, we (as listeners following the music in real-time) can never be certain whether a double return articulates a true recapitulatory beginning or a false one. When encountering a thematic return, listeners will most likely subscribe to the hypothesis that this is the true recapitulatory beginning, provided that this hypothesis is supported by other, more contextual criteria (such as adequate harmonic preparation, the temporal location of the thematic return, and a sufficient degree of resemblance to the original P). This might encourage analysts to attend more to surface-oriented contextual features that help to clarify or communicate formal functions.

(2) Although the false-recapitulation concept is ubiquitous in analytical writings, some authors adopt a comparatively critical stance toward it, as for instance Hepokoski and Darcy, who emphasize its "hermeneutic weakness:"[74] "It is a mere label, claiming nothing more than the registering of a momentary personal deception. By itself it explains nothing about the piece."[75] The authors further argue that a clear-cut either-or decision as to whether or not a tonic-key restatement of P represents a false recapitulation is beside the point: "The reality is that we may confront a potential false-recapitulation effect in varying strengths, along a continuum […]."[76] What is remarkable about Sonata Theory's understanding of the false recapitulation is that Hepokoski and Darcy explicitly define the surprise effect imparted by this device in intentional terms, namely as "the degree to which Haydn [or, more generally, any composer] *intended* the listener to be misled with such a tonic-P-statement."[77]

To understand why intentions are of such importance to the notion of false recapitulation, or to falseness in music in general, it is worthwhile to remind ourselves that to speak of a musical event as a false recapitulation is not a literal statement about a musical fact. Rather, it is an interpretation in terms of a linguistic ('illocutionary') speech act[78]: If a given event does not match listeners' expectations as based on their prior knowledge of conventional patterns, i. e., if listeners are not able to make sense of that event right away, they most probably will not reject it as meaningless, arbitrary, or flawed; rather, they may assume some larger underlying communicative purpose on the part of the composer, based on Paul Grice's "cooperative principle"[79]; this assumption prompts them to make

---

72 Maus 1999.

73 Levy 1982.

74 Hepokoski / Darcy 2006, 224. On the limited value of surprise as a general explanatory concept, see David Rosen 1996, 266.

75 Hepokoski / Darcy 2006, 244.

76 Ibid.

77 Ibid., 223 (my emphasis).

78 The following train of thoughts is inspired by a comparable discussion of Haydn's wit as manifest in the String Quartet op. 33 No. 2/iv proposed by Justin London (1996).

79 Grice 1975, 45ff.

special efforts to reconstruct this purpose. In the case of the false recapitulation, listeners (and modern analysts) ascribe to it specific intentions, namely those of a deliberate play on listening expectations on the part of the composer, rather than interpreting the premature return of the primary theme in the tonic (or another closely related) key as a compositional flaw.[80]

This suggests that to understand a particular double-return instance in the development section only as a deviation from a convention is insufficient to explain why this event might be regarded as witty. Identifying nonconformance is only the first step; it is also necessary to recognize that by interpreting this event as false, we implicitly treat the composer as a speaker who intends to communicate through a given composition (the utterance), assuming that the musical means he or she chooses are *meaningful* and *coherent* and can indeed be fully understood by listeners. In Hoyt's words, "not only must the listeners be fooled, they must believe that they were *intended* to be fooled."[81] In so doing, listeners implicitly invoke the underlying metaphor of music as a means of communication and hence (what cognitive metaphor theory calls) the "source domain" of music as language.[82]

If we accept authorial intentionality as lying at the heart of the false-recapitulation concept, we can now move on to consider the subsequent question of whether it is indeed plausible to think of a double return in the development as an intentional play on expectations, based on the extent to which a number of essential criteria (or indicators) are fulfilled in a given corpus of works.

When trying to determine the intended (as opposed to the hypothetical "actual") strength of the false-recapitulation effect, Hepokoski and Darcy outline seven questions, one of which is the issue of "[h]ow literal the tonic-P-reference [is]."[83] Clearly, the deceptive effect associated with the false recapitulation increases when the P-restatement closely resembles the original statement of the primary theme found at the very beginning of the movement. If the second-half occurrence of P differs markedly from its original statement in the exposition, the authors consider it unlikely that the composer intended his audience to be misled by this device. Instead, it is reasonable to think of the thematic variation at hand as a subtle hint to the listener that the primary-theme instance in question is not meant to act as a formal marker of long-range significance. In this regard, in a number of cases Hepokoski and Darcy observe striking differences between the primary themes launching the supposedly false recapitulations and the original P-instances found at the movements' outsets.[84] Based on this observation, and in accordance with their

---

80 Here, the identity of the composer plays a crucial role: With a lesser-known composer, analysts are more inclined to consider this a flaw; with Haydn, we assume that some larger issue is at stake here. In so doing, we rely on and continue to confirm our imagine of Haydn as the ever experimenting scientists who manipulated sounds and observed the resulting listening reactions (Griesinger 1810, 24f.).

81 Hoyt 1999, 30.

82 On the concept of source domain, see for instance Lakoff 2008, 276.

83 Hepokoski/Darcy 2006, 225.

84 Ibid.

idea of the false-recapitulation effect as lying on a continuous spectrum, they argue that the sense of a false recapitulation is considerably weakened in such cases.

For instance, in the final movement of his String Quartet op. 50 No. 1, Haydn introduces subtle variations in the P-restatement (mm. 109ff.) in the development section. Heralded by an extended home-key dominant (mm. 103–108), the primary theme is embedded within a contrapuntal texture from m. 112 onwards, which may be seen as an indication that this thematic return is not intended as a deliberate deception on the part of the composer. In his Symphony No. 71,i (ca. 1779/80), Haydn chooses two different main themes to articulate the return to the tonic key. While the first one (mm. 121ff.), prepared by an extended pedal on V/vi (mm. 106–120), presents a P-function that has not been heard before, the second instance (mm. 169ff.) is derived from the opening of the movement. These differences between the exposition's P and the subsequent tonic-key P-function may be viewed as an indication that Haydn did not want his listeners to hear the latter as the *initium* of a true recapitulation. In other words, it is plausible to assume that had Haydn intended to convey a false-recapitulation effect, he would have used a version of P more similar to the P-statement heard at the movement's outset.

In addition to the question of the literalness of P, a second important criterion concerns how the supposedly false tonic-P return is prepared, in particular whether or not it is preceded by "its 'proper' dominant."[85] In this regard, it is interesting to note that some of the instances (around 1770) commonly characterized as false recapitulations lack such a dominant. They are preceded instead by so-called 'bifocal retransitions' (V/vi–I).[86] This is a remarkable finding, as Haydn had begun to experiment with bifocal retransitions to *true* recapitulatory beginnings in the tonic key some time before (in the late 1760s), and he was not alone in employing this device; in fact, numerous composers of the time made use of this practice.[87] There was nothing deficient about this strategy, which was derived from, and at the same time transformed, an earlier baroque practice; it would not necessitate compensatory moves later in the movement.[88] However, despite its frequency of occurrence, the bifocal retransition is certainly not the most widely used option; rather, it is what Hepokoski and Darcy call a "second-level default."[89] If this is the case, then one might raise the question why Haydn would employ this option in preparation of false recapitulations (in the tonic and subdominant keys)[90] instead of clearly communicating the recapitulatory function of the double return by choosing the first-level option (V/I). In other words, had Haydn wanted his audience to hear these re-

85 Ibid., 224.

86 LaRue's term (1992).

87 Neuwirth 2009 and Neuwirth (in preparation).

88 Ibid.

89 Hepokoski/Darcy 2006, 10.

90 See Haydn's Symphony Nos. 42,i (mm. 87–89), 43,i (mm. 105–113), and 46,i (mm. 66–70), as well as his String Quartets opp. 20 No. 1,i (mm. 40–43) and 54 Nr. 3,iv (mm. 83–95). An early example of a bifocal hinge to a mere tonal 'medial return' can be found in Haydn's Symphony No. 31,i (*Hornsignal*, from 1765; mm. 69–71). False recapitulations in the key of IV that are prepared by a bifocal progression are found in Haydn's String Quartet op. 20 No. 4,i (mm. 203–206) and in his Symphonies Nos. 51,i (mm. 107–109) and 80,iv (mm. 159–168).

turns as true recapitulations, he would presumably have chosen the home-key dominant prior to the reentrance of P, since this would have ensured a higher degree of similarity between false and true recapitulations[91]—and this type of question is not only relevant for the analyst but also for the listener who is forced to make sense of Haydn's formal strategy in real-time.

While Hepokoski and Darcy conceive of these two changes—altering the P-restatement and preparing it by sonorities other than V/I—as possible means of weakening the false-recapitulation effect, they do not go so far as to deny the existence of this effect entirely. In this regard, Ethan Haimo has gone one step further: He seeks to provide strong evidence against the idea that a double return occurring in the development conveys a deceptive effect at all. Haimo argues that double returns that supposedly turn out to be false differ in one important respect from true recapitulatory beginnings, namely with regard to their 'temporal location': Typically, these double returns appear somewhat too early to be mistaken for the true recapitulation.[92] As Haimo's statistical analysis of a selection of Haydn's symphonic movements reveals, recapitulations (on average) enter once the development has reached about 60% the size of the exposition. By contrast, in Haimo's sample, the distance between the beginning of the development and the entrance of the false recapitulation represents only (at a maximum) 45.5% of the exposition's length.[93] This suggests the intentional argument that had Haydn indeed intended to deceive his listeners with regard to the recapitulatory status of a double return, he would have opted to delay the entrance of P until the development had reached the more normative length of about 60% of the exposition. As logical as this argument may seem, it is doubtful whether Haydn could expect his listeners to be near-perfect statisticians, capable not only of calculating proportional differences in real-time but also of drawing on a mental representation of the entire sample on which Haimo's statistical analysis is based (see the above premise concerning internalized probabilities).[94] This premise is all the more problematic when one considers that especially in small-scale works, absolute differences may be much less salient than relative percentages might suggest: In Haydn's Symphony No. 55,i (the focal point of Haimo's argument), the value of 45% before the entrance of P (m. 97) is in fact misleading, since there is only an absolute difference of nine measures to the "normative" percentage value of 60%. It is difficult to imagine that listeners were able to discern such a small difference, even listeners of the "educated" type to which Rosen alludes.[95] Even if they were, it is highly unlikely that Haydn would take the risk of jeopardizing the success of his formal game by exclusively relying on such a small difference in the temporal location of P.[96]

---

91 It is inconsistent to claim that these bifocal retransitions were unconventional (weak) preparations of the recapitulation and at the same to assert that Haydn used them in preparation of main theme returns that he wanted to be understood as real recapitulations; see Neuwirth 2009.

92 See Haimo 1995, 106–113. See also Rosen 1988, 280f.: "The false reprise is a false repose: both the way it enters and the way it disappears are conceived as shocks, as both come too soon."

93 Haimo 1995, 110.

94 Cf. also the remarks by David Rosen (1996, 267) on the knowledge of the "standard operating procedure" in Mozart's piano concertos that the composer could expect from his audience.

95 Cf. also Hoyt 1999, 83.

The issue of temporal location has previously been invoked by theorists who endorse, rather than reject, the concept of false recapitulation. For them, a genuinely false recapitulation is to be distinguished from another technique, the 'early medial return'[97] of the main theme and the tonic key shortly after the beginning of the development (deriving, according to Rosen, from the mid-eighteenth century opera aria[98]). It is argued that this technique did not have the potential of misleading the listener with regard to its formal status, as it occurred too early in the movement to be confused with a true recapitulatory beginning.[99] Nevertheless, the early medial return and the false recapitulation are not construed as entirely distinct; rather, they are said to be genetically related, the early medial return being the precursor of the false recapitulation: By delaying the appearance of the early medial return and concomitantly enlarging the developmental space preceding the moment of tonal stability, a deceptive effect could be created. As Webster hypothesizes, "[t]he 'false reprise' may well have originated in part as a displacement of the immediate reprise 'forward', as it were, into the body of the development."[100]

This statement raises interesting questions, such as the following: When did listeners begin to perceive a double return situated in the development section as a play on listener expectations? When was the ternary model first perceived as a norm that enabled composers to play with the idea of the recapitulation? These questions lead us to a critical evaluation of the historical premises underlying the concept of false recapitulation.

(3) Since the play upon a norm or convention is by definition only possible once this norm has been firmly established, it is widely assumed that eighteenth-century composers (and Haydn in particular) began to experiment with the double return from about 1770 onwards, i.e., starting around the time when the ternary model of the sonata form was gradually superseding the older binary model and when the moment of recapitulation began to acquire the character of an emphatic event.[101] Compositions after 1770 are more likely to be classified as 'sonata form with false recapitulations,' while compositions written before 1770 are more likely to be referred to as featuring an 'early medial return.'

As straightforward as this distinction may seem, the criteria for labeling a given piece as a sonata form with false recapitulation or as featuring an early medial return have not been employed in any systematic manner, as Hoyt persuasively demonstrates.[102] Likewise, Burstein observes that "[t]here is no universal agreement regarding how to distin-

---

96 One might argue that the formal strategy employed in Haydn's Symphony No. 55,i is, if anything, a play upon the convention of *continuing* the early medial return (see note 97), rather than on the recapitulation convention.

97 This is my coinage, based on Hoyt's term 'medial return' (see note 31).

98 Rosen 1988, 37, 39.

99 When this technique was revived much later in some works by Brahms (e.g., his Symphony No. 4,i), it is again not ascribed any deceptive potential but rather a non-structural, or non-dramatic, relaxation of tension accompanied by the introduction of lyricism.

100 Webster 1986, 128. See also Webster 1981, 526f. and Rosen 1988, 156–161.

101 As Rosen observes, "[o]nly gradually did composers come to believe that avoiding the tonic altogether before thematic resolution would be more effective; only during the 1770s did they firmly equate the final return of the tonic and the return of the main theme […]" (1988, 157).

102 Hoyt 1999, 159ff.

guish between 'premature recapitulations' and 'false recapitulations'."[103] In the absence of such criteria, analytical categorizations have often been biased with respect to chronological issues; the fact that the two groups may have some important defining features in common has largely been overlooked. The resulting inconsistencies in the classifications of relevant examples can even be found within the context of the same approach.

For instance, Bonds classifies the second movement of Haydn's E-flat major Symphony No. 11 (1760/61) as a hybrid between an early medial return and a false recapitulation, whereas he interprets the double medial return in the D-major Symphony No. 42,i (1771) as an unequivocal instance of a false recapitulation.[104] This might seem to be due to a somewhat inconsistent decision, because in the second case the development (up to the moment of the medial return) is nine measures shorter than in the first one (7 vs. 16 mm.); in addition, the 7-m. "development section" in the D-Major Symphony consists of nothing but 'standing on the dominant' of the relative minor (mm. 82–88), which no doubt represents a second-level default for the preparation of the recapitulatory double return (m. 89).[105]

Similarly, the medial double return in Haydn's C-major Symphony No. 37,i (1757/58) is classified as a 'disjunct recapitulation' (a variant of the early medial return to be found in binary sonata-form movements), whereas the analogue event in Haydn's Symphony No. 38,iv (1767) is described as a hybrid between an early medial return and a false recapitulation, despite the fact that there is no essential difference between these two situations: Both in Symphony No. 37 (mm. 82 ff.) and in No. 38 (mm. 75 ff.), the medial double return appears after roughly the same stretch of time (13 and 12 mm., respectively), during which a dominant sonority is prolonged. The fact that these dominant functions appear in different tonal contexts (in No. 38 a V/vi, in No. 37 a V/I) does not seem to justify such a categorical distinction: to perceive a development "proper" in No. 38 and thus to hear the subsequent double return as the initial event of a true recapitulation lacks any solid foundation.[106]

Table 1 distinguishes between early medial returns and false recapitulations in Haydn's *oeuvre*, providing additional chronological and genre-specific information. This table suggests that instances of false recapitulations can predominantly be found in two (highly sophisticated) genres, the symphony and the string quartet. The vast majority of medial returns can no doubt be found in the representative genre of the symphony, especially in the 1760s and around 1770. Early medial returns (or disjunct recapitulations) are by no means peculiar to Haydn; they can also appear in many works of his contemporaries, as will be shown below.

---

103 Burstein 2011a, 8.

104 Bonds 1988, 240 and 312–315.

105 See Neuwirth 2009.

106 More difficult to assess might be the first movement of Haydn's D-major String Quartet op. 20 No. 4. The section between the double repeat sign and the first entrance of the main theme in the tonic key (m. 133) comprises 20 measures. Yet this section seems to contradict the notion of development, since only dominant pedal points are used (V/ii and V/iii).

Chronology	Symphony	String Quartet	Keyboard Sonata	Keyboard Trio
-1761	2/i, 2/ii, 4/i, 5/iv, 10/i, *11/ii, 15/i, 15/ii, 18/i, 32/i, 33/i, 37/i, A/i, B/iv	1,0/i, 1,2/i		
~1761 – ~1763	3/i, 14/ii, 14/iv, 17/i, 19/i, 19/ii, 25/i, 36/i, 36/ii, 36/iv, 9/ii			
1761-1765	7/iv, 8/ii, 13/iv, *22/ii, 29/i, 29/iv, 31/i, 31/ii, 72/i		13/i, 3/ii 1/ii, 6/i	38/iii, *41/i, 36/i, 2/i, 12/i
1766-1772	38/iv, *41/i, *42/i, *43/i, *46/i, *48/iv	9,2/iv, 17,1/i, 17,6/iv, *20,4/i, *20,1/i		2/i
1772-1781	51/i, *55/i, 65/i, 65/ii, 70/i, *71/i	33,5/i, 42/i		*12/i
1780s		*50,1/iv, 54,3/iv		
1790s	*91/i	*77,1/i		

Table 1. Early medial returns and false recapitulations (in the tonic key); genre-specific and chronological distinctions in Haydn's *oeuvre*. This table is largely based on Schmidt-Thieme (2000, 136); Schmidt-Thieme's table is a corrected version of the list reproduced by Bonds (1988, 308–309, and 316). * = "false recapitulation."

It is an open, and perhaps even speculative, question where Haydn first encountered the strategy of the (early) medial return. One possibility is that he has learned it at a very young age during his first stay in Vienna from Georg Reutter, who made use of this strategy in the *Intrada* to his *Servizio di Tavola* (1757). However, it is also possible that, in using this strategy, Haydn simply followed a widespread Viennese (or, more generally, Austrian) convention, as exemplified by works of Wagenseil, Schlöger, L. Mozart, W. A. Mozart, Adlgasser, Dittersdorf, Vanhal, Asplmayer, and Salieri written during the second half of the eighteenth century.[107] But even this suggestive claim may be contested in view of the fact that the (early) medial return was also customized in the 1760s and early 1770s at various other influential European centers, e. g., in Mannheim (W. Cramer and Richter),[108] Berlin and other northern German centers (C. P. E. Bach, J. A. Benda, Neefe, and Wolf),[109] London (J. Chr. Bach),[110] and Paris (Schobert).[111]

107 Asplmayer, Symphony in F,i; L. Mozart, F-major March; Allegro No. 1 in C; Allegro No. 2 in C; and W. A. Mozart, Allegro K. 6,i; Keyboard Sonata K. 283,ii; Salieri, Symphony in D,i; Dittersdorf, Symphony in E-flat,i; Adlgasser, Symphony in A,i. For further examples by Wagenseil, Schlöger, and Vanhal, see Neuwirth, in preparation.

108 W. Cramer, String Trio op. 3 No. 2,i; Fr. X. Richter, String Quartet op. 5 No. 3,i.

109 C. P. E. Bach, Prussian Sonata No. 5,i; Württemberg Sonata No. 4,i; J. A. Benda, Keyboard Sonatas Nos. 1,iii, 8,iii, and 16,i, Symphony No. 3,i in C (-1762); C. G. Neefe, Sonatas Nos. 10,i and 11,i; E. W. Wolf, Sonata No. 2,iii in G (1774).

Rosen's claim that "during the 1770s [composers] did […] firmly equate the final return of the tonic and the return of the main theme"[112] and abandoned the practice of the early medial return entirely downplays the fact that even after 1770, tonic returns in the development section can be found in a not insignificant number of compositions; in other words, this claim does not stand up to empirical scrutiny.[113] Not only Haydn, but also some of his Viennese contemporaries, including W. A. Mozart and Koželuch, employed this device at least occasionally in works from the 1780s and 90s.[114] Since at present we know only a small portion of the entire output in this period—the "tip of the iceberg," as it were—, it is by no means unlikely that much more works featuring such (early or late) medial double returns by supposedly minor composers can be discovered. Until then, the claim that this practice virtually disappeared in the classical period is premature, as it is based on a small and largely unrepresentative sample (mainly including only the works of Haydn and Mozart). Nevertheless, based on these preliminary results we can draw at least the tentative conclusion that listeners of the time were not necessarily surprised when a medial return did not turn out to be the beginning of the true recapitulation (that is, one followed by the remaining elements of a complete thematic rotation in the tonic key).

The following two analyses, one of a movement by Haydn, the other of a movement by Koželuch, will demonstrate that developmental and recapitulatory procedures may at times be much more closely intertwined than the modern paradigm of sonata form suggests, which considers the double return as a reliable demarcator of the recapitulatory action space. If modern analysts are facing serious problems, even after hours of deep reflection, in trying to make sense of the formal procedures realized in these movements, how then could an eighteenth-century audience resolve the formal puzzle at hand under the constraints of real-time perception?

The fourth movement of Haydn's E-major String Quartet, op. 54 No. 3 (1788) is a fitting (late) example of the 'early medial return' strategy. Announced by a V/vi dominant prolongation (mm. 83–94), the primary theme enters, supposedly prematurely, in the tonic key (mm. 95ff.) shortly after the beginning of the development section. The issue of where the recapitulation begins is controversial. One may argue that the true recapitulation is articulated by the return of the primary-theme-based subordinate theme[115]

---

110 J. Chr. Bach, Keyboard Sonata op. 17 No. 5,i.

111 Schobert, Trios op. 16 No. 1/i and op. 16 No. 4/i; Violin Sonata op. 14 No. 5/i.

112 Rosen 1988, 157.

113 Nevertheless, Rosen seems to have been aware of the anachronistic implications inherent in this type of historical assessment: "Evidently, composers sometimes had a certain shyness about bringing back the opening bars, particularly when they had already been played a second time at the dominant. This psychological resistance to the modern conception of recapitulation as a simultaneous return to the tonic and the first theme is significant: the second half of the sonata is never really considered as two absolutely separate sections until the nineteenth century" (ibid., 150f.).

114 Koželuch, Keyboard Sonatas opp. 1 No. 1/i, 1 No. 2/iii, and 8 No. 2/i, as well as his Symphonies I:3/iv, I:4/iv, and I:6/iv; Mozart, Violin Sonatas K. 377/i and K. 526/i.

115 Note that the P-based S begins on a different (relative) scale degree (1 instead of 3) but retains the relative intervallic distances (cf. mm. 1ff. ~ 45ff.), thus guaranteeing recognizability.

(mm. 165ff. ~ mm. 45ff.) in the sonata form's second half, subsequent to a long passage that assumes the character of a developmental retransition following a converging I:HC (m. 134). However, this interpretation is by no means unproblematic, since it implies the omission not only of the primary theme but of the ensuing transition section (mm. 16–44) as well. The considerable proportional imbalance between exposition and recapitulation (82 vs. 54 mm.) resulting from these alterations is rather difficult to explain.

No matter whether the early medial return in Haydn's op. 54 is interpreted as a deliberate deviation from a contemporary norm or as the revitalization of an older convention, one can argue that Haydn's choice of omitting the primary theme at the start of the recapitulatory rotation was *motivated* by Haydn's use of the primary theme (which is altered from the sixth measure onwards) earlier in the movement.[116] One might even go so far as to take this occurrence of the primary theme as being the "real" *initium* of the recapitulatory rotation, and accordingly interpret the digressions that follow it (mm. 101–165) as being a greatly expanded interpolation interrupting the expected course of events, before the music returns to passages already familiar from the exposition.[117] This suggests that the development section is partly recapitulatory in nature, thus blurring the distinction between the two action zones. That Haydn could "[e]xcise from the true recapitulation some or all of the material already presented in the false recapitulation"[118] clearly underscores the genuinely recapitulatory nature of the supposedly false or premature tonic return.

Another instructive example of the formal ambiguities resulting against the backdrop of the modern paradigm of sonata form can be found in the second half of the opening movement of Leopold Koželuch's Keyboard Sonata, op. 1 No. 1 from 1780 (see Fig. 1). On a stylistic level, this movement is not at all complex (it is written in a straightforward symphonic style); however, it creates notable formal-functional complexity by combining various formal strategies in a manner that is—at least from the perspective of modern sonata-form theories—highly unusual, rendering a clear-cut distinction between developmental and recapitulatory action spaces virtually impossible, both for analysts and listeners (modern and contemporaneous alike).

The second half of this movement commences with the initial 12 mm. (the A section) of the primary theme[119] in the key of V—a common strategy employed in numerous pieces from the 1760s and early 1770s that suggests a formal design sometimes referred to as 'counter-exposition.'[120] Following the (potential) counter-expositional model, Koželuch

116 Nevertheless, this movement is clearly not an example of a 'disjunct recapitulation' because what appears in m. 165 is not the consequent of the main theme (the antecedent of which is restated shortly after the beginning of the recapitulation) but rather a retransposed version of the P-based S, articulating the moment of crux. In many ways this strategy produces a formal design that strongly resembles a subtype of binary sonata form, one that features an early medial return.

117 After the medial return, a digression to various keys takes place, eventually leading to C♯-minor (the key of vi) in m. 126. By using a 7-6 sequence starting in m. 129, Haydn modulates back to the home key, which is eventually confirmed by a converging I:HC. Although the home key is cast into doubt in m. 142, one comes safely back to it in m. 162 (V/I).

118 See, according to Bonds, Haydn's Symphonies Nos. 36,iv, 41,i, 46,i, 71,i, 91,i; String Quartets opp. 20 No. 4,i and 54 No. 3,iv (1988, 346).

119 This 12-m. group forms the A section of a larger ABA' structure.

120 Cf. Somfai 1995, 283.

repeats the next unit from the exposition, the linking phrase mm. 13–16 (B), which is altered only slightly at the end, so that it concludes in m. 78 with a weak imperfect authentic cadence (rather than a HC, as in the exposition). It is notable that this cadence confirms the tonic rather than the dominant, thus enabling the subsequent entrance of the primary theme's A' section (mm. 79ff. ~ 17–20 ~ 1–4) in the tonic as well, in the manner of an early medial return. This return might not be particularly noteworthy; from a historical perspective, however, the use of this device in a composition from the 1780s is commonly be seen as a rarity, since this practice is assumed to have been largely abandoned in the course of the early 1770s.

Exposition (mm. 1–62)

Measures	1–12	13–16	17–20	21–33	34–45$_e$	45–53$_e$	53–62
Functions	P: A	B	A'	TR^{1+2}	S (P-based)	S-cont.	C
Tonality	I:PAC	I:HC	no cad.	V:HC	V:PAC$_{dec.}$	V:PAC^1	V:PAC^2
Dynamics	piano			forte	piano	forte	p – f

Second half (mm. 63–150)

Measures	63–74	75–78	79–82	83–96		122–133$_e$	133–141$_e$	141–150
Material	P: A	B	A'	TR^1		S (P-based)	S-cont.	C
Tonality	V:PAC	I:IAC	no cad.	IV:V$^6_5$/ ♭III		I:PAC$_{dec.}$	I:PAC^1	I:PAC^2
Dynamics	piano			f – p		piano	forte	p – f

97–100	101–108	109–121
P: A	Fantasia Style	TR^2
♭III:	⟶ vi	vi ⟶ i:HC

Figure 1: Analytical graph of the formal structure of Leopold Koželuch's Keyboard Sonata op. 1 No. 1/i, exposition (mm. 1–62)

One way of legitimizing Koželuch's seemingly eccentric decisions might be to invoke the notion of dialogue introduced to the scholarly discourse by Hepokoski and Darcy: We could assume that Koželuch intended to make a reference to (or to revitalize) a by-then obsolete formal practice.[121] However, beyond assumptions regarding the composer's intentions, we should not overlook the fact that within the *intra-opus* context of this particular movement, the simultaneous return of primary theme and home key is (at this point in the form) the logical continuation of the rotational pattern that was launched at the beginning of the development section in the key of V (mm. 79–82 are analogous to both mm. 1–4 and mm. 17–20). This strictly rotational reading receives additional support from the fact that the subsequent passage corresponds to the next unit in the reference

---

121 For instance, the early return to P in V shortly after the beginning of the development in Haydn's String Quartet op. 33 No. 6,i (1781) is understood as a dialogue with the older convention of the early medial return and the later off-tonic recapitulation, see Hepokoski/Darcy 2006, 276f.

pattern, the exposition transition, which is transposed down two fifths (thus appearing in the key of IV; mm. 83–86 ~ 21–24)—a typically recapitulatory procedure that normally ensures the re-appearance of the secondary theme in the home key. However, contrary to what one might expect at this point, the secondary theme does not materialize. In fact, after the transposed restatement of the transition section, Koželuch breaks the rotational pattern by postponing the entrance of the P-based secondary theme, which occurs as late as m. 122. Instead of strictly adhering to the harmonic pattern, Koželuch sequences the four transitional measures down a third (mm. 87–90), leading to the next four-measure group (mm. 91–94), one that has a strong D-minor implication. However, this implication is eventually denied, with c♯ reinterpreted as d♭ thus setting the stage for the key of ♭III (A-flat major). While the turn to the flattened mediant key is not problematic in itself, the fact that its entrance is articulated by means of the first phrase (A) of the primary theme (mm. 97–100) poses problems of interpretation, since it might temporarily give the impression of an off-tonic recapitulation. However, the music quickly abandons the thematic parallelism to P: An 8-m. *fantasia*-like digression (with no counterpart in the exposition or elsewhere in the movement) moves away from ♭III, eventually fulfilling the previous promise of D minor (the relative minor of the home key), again after a fully-diminished seventh chord (m. 108 ~ 91 or 93).

Koželuch resumes the rotational pattern just at the point he had left it, restating the second half of the transition (TR^2; mm. 25ff.)—its first half was heard in mm. 83ff. after the early medial return—first in D minor, then in B-flat major, and eventually in the home key. TR^2 finally gives way to the moment of crux (mm. 113ff.) and, subsequently, to the reentrance of S: After a Phrygian half cadence in the tonic minor, the long-delayed S ultimately reappears in the tonic major key (mm. 122ff.). It is only at this moment that the tonic key is firmly re-established for the first time. Since S is initially based on the primary theme (m. 122 ~ 1), its entrance could at least briefly seem to articulate the recapitulation proper.

Drawing on traditional sonata-form criteria, one can analyze the formal design realized in this movement in one of the following ways: (1) The fact that the (second) return to the tonic coincides with S, rather than P, could suggest a reading of this movement as a binary sonata form. The (by-then) obsolete early medial return represents a moment of tonal stability *within* the action space of the development section. (2) The fact that S is preceded by a (partial) return of P in ♭III could provoke an interpretation of this movement as one featuring an off-tonic recapitulation. As in the previous reading, the early medial return does not have any recapitulatory implications and belongs conceptually to the development. (3) One could also view the early medial return, as the *only* restatement of P in the tonic during the entire second half, as the initial event of an extensive recapitulation that is internally expanded by tonal digressions to IV and ♭III—a reading that faces obvious conceptual problems. (4) Finally, one can alternatively consider the entire second half, starting with P in V, to be a somewhat expanded replication of the referential pattern established in the exposition, meaning that the question of where the boundary between the development and the recapitulation lies would be left unresolved. Since recapitulatory and developmental procedures are so closely intertwined in this movement, our analytical basic concepts of recapitulation and development become virtually

meaningless. In a way, the entire second half of the movement can be referred to as a greatly expanded (thematic) recapitulation, realizing a highly circular design that is characterized by constant motion away from and returning to the tonic key and/or the main theme. Although the methods by which this movement blends principles of ternary and binary sonata forms create ambiguity for the *modern* analyst and/or listener, Koželuch's use of the strategies described is arguably entirely consistent with *eighteenth-century* theoretical expectations, as we shall see in the next section.

3. The 'double medial return' from the perspective of eighteenth-century writings

The preceding discussion points to the necessity of revealing the specific structural functions of the medial return based on the analysis of individual works, rather than simply relegating it to the status of a false recapitulation—a mode of explanation that leaves any genuinely structural motivation of this formal strategy out of account. Moreover, it should have become clear by now that interpreting the double return within the development section as a false recapitulation is based on anachronistic assumptions deriving from the modern paradigm of sonata form. If this is so, a historically adequate reinterpretation of this phenomenon is required, one that receives additional support from music-theoretical treatises and composition manuals from the eighteenth- and early nineteenth centuries. As mentioned earlier, historical writings constitute a second source that informs a modern analyst's attempts at a hypothetical reconstruction of the expectations of eighteenth-century listeners.[122]

Unlike numerous twentieth-century theorists, some writers of the eighteenth century, including Riepel, Löhlein, Kirnberger, Koch, and Gervasoni, demanded the use of the main theme in conjunction with the main key in the section we now call development. This device was seen to provide a tonal anchor or point of reference to listeners, in addition to helping them to recall the main idea of the piece. What nowadays is commonly seen as a witty deception of the listener—a seemingly premature resolution of a large-scale dissonance—is, in terms of historical theory, viewed as an effective means of guiding the listener along a complex tonal journey. The following quotations from eighteenth- and early nineteenth-century sources, in which historical theorists reveal themselves as implicit psychologists, serve to substantiate this argument.

As is widely known, the third volume (1793) of Heinrich Christoph Koch's *Versuch einer Anleitung zur Composition*, which offers an account of what we would nowadays refer to as sonata form, explicitly mentions the option of returning to the tonic key within the first "Hauptperioden" of the second part ("zweyter Theil"). In no way does Koch refer to a formal anomaly or an event that would require special explanation by invoking the composer's intent to deceive listening expectations. On the contrary, he calls this

---

122 Although statements about how people actually listened in the eighteenth century are of course highly speculative, such statements can differ from one another with respect to their plausibility, based on the interpretation of a number of indicators that inform us, more or less directly, about listening behavior and expectations.

method of designing the development section "the first and most usual construction," which he describes as follows:

> [...] the first period of the second section begins in the key of the fifth with the theme, occasionally also with another main melodic idea, either note for note, in inversion, or also with other more or less considerable alteration. After that it either modulates back in 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.[123]

In Koch scholarship, it is rightly maintained that Koch's description does not reflect the compositional practice of his time, but rather refers to an older style practiced in the 1760s and 1770s (as is evident from the repertoire that Koch draws upon). Therefore it is not surprising to find descriptions of the 'early medial return' technique that in many respects anticipate Koch's own account, as for instance in Georg Simon Löhlein's *Clavier-Schule* from 1765:

> The main key must be impressed upon the ear right from the start through neighboring chords. Once this has happened, one turns to the key of the fifth g [...]. This is accomplished by its raised seventh f♯. In this very key one returns to the [main] theme, and dwells on that key and then brings a close [...]. After that, one usually returns to the main key c by the naturalized f♯ [...], and moves on, via g♯, to the key of the sixth, A minor [...] Now one returns to the main key [...], and touches on the flattened seventh B*b* the fourth of the main key F major [...], in order to clarify the main key to the ear.[124]

An even earlier approach can be found in Joseph Riepel's *Grundregeln zur Tonordnung insgemein* from 1755, in which key relations are conceptualized in terms of social hierarchy. Riepel observed "that the *Meyer* or the main tone C often reappears in the middle, as if it has to give new commands all the time. In other words, it [the tonic] must disappear from neither the eyes nor the ears. Everything turns and twists around it, like

---

123 "[...] das er mit dem Thema, zuweilen auch mit einem andern melodischen Haupttheile, und zwar entweder von Note zu Note, oder in verkehrter Bewegung, oder auch mit andern mehr oder minder beträchtlichen Abänderungen in der Tonart der Quinte angefangen wird, nach welchem entweder vermittelst eines andern melodischen Theils die Modulation zurück in den Hauptton geführt, und von diesem in die weiche Tonart der Sexte oder auch in die weiche Tonart der Secunde oder Terz geleitet wird [...]" (Koch 1793, 307 f.; transl. in Hoyt 1999, 59).

124 Löhlein 1765/1782, 182: "Diese Haupttonart muß dem Gehöre, gleich am Anfange, mit ihren Nebenaccorden wohl eingeprägt werden [...]. Wenn dieses geschehen [183:] ist, so wendet man sich in die Quinte g [...]. Dieses geschicht [...] durch dessen erhöhete Septime ♯f. In dieser Tonart nimmt man wieder das Thema vor, und hält sich auch wieder etwas darinnen auf, und macht sodann einen Schluß [...]. Hierauf wendet man sich ordentlicher weise, durch das wieder hergestellte fis, in die Haupttonart c zurück [...], und gehet sogleich, durch Vermittelung des ♯g, in Sextam modi a Moll [...]. Oder man kann auch statt in diese, in die Terz e Moll gehen. Man kann auch hier einen ordentlichen Schluß machen, oder ihn auch, wenn man sich nicht lange aufhalten will, übergehen. Nunmehro wendet man sich wieder in die Haupttonart [...], und berühret durch die erniedrigte Septime *b*h, die Quarte der Haupttonart f Dur [...], um dem Gehöre wieder die Haupttonart c zu schärfen" (my translation).

the cat circling the hot broth. Through him one may immediately reach those under his power […]."[125]

Whereas Löhlein and Koch provided a more descriptive approach, Johann Philipp Kirnberger (in his treatise *Die Kunst des reinen Satzes in der Musik*) points out the underlying cognitive rationale for choosing a medial return when realizing a modulatory process. In keeping with Riepel's account, he advises the composer to take care that the listener does not lose sight of the main key during modulations to various subordinate keys:

> In […] modulation it is a general rule to proceed in such a way that the main key, in which the piece begins and ends, is never completely erased. Therefore one should modulate to another key only after the ear is almost satiated with the main key, and these secondary keys should not mask the main key to the extent that it is completely forgotten. Thus one must always stay in its neighborhood, so to speak, and renew its feeling from time to time. Wherever this is neglected, it is difficult to preserve harmonic unity.[126]

Furthermore, Carlo Gervasoni, whom Rosen tellingly dismissed as a "reactionary theorist"[127], emphasizes not only the tonal component but also the cognitive function associated with thematic repetition. As late as 1800, he asserts in his *La Scuola della Musica* that "after such a repeat sign, the main mode reenters, thus impelling [one] to recognize the motif identified with the principal mode; such reiterations admirably serve to strengthen the expression and to recall the opening idea of this sonata."[128]

Finally, there is Anton Reicha's 1826 description of the development section of sonata-form movements: "The key of D (the tonic) and the key of A major must here only be found in passing. The first, because it must predominate in the second section [the recapitulation], the second, because it has been used extensively in the first part [the exposition]."[129] Although Reicha permits a return to the tonic after the double repeat, he recommends that this should be a subordinate and brief event. One might hypothesize

---

125 Riepel 1755, 67: "[…] daß der *Meyer* oder *Hauptton* C. auch in der Mitte wieder oft vorkommt; gleichsam als wollte er immer neue Befehle oder Berichte ertheilen. Mit einem Wort, *er muß weder aus den Augen noch aus den Ohren gelassen werden*. Alles windet und wendet sich um ihn herum, als wie die Katze um den heissen Brey. Durch ihn kann man den Augenblick zu allen seinen Untergebenen gelangen […]" (My translation is based on those by Hoyt 1999, 301, and Brover-Lubovsky 2005, 19). See also Koch 1787, 169–171.

126 Kirnberger 1771, 107 (transl. in Beach/Thym 1982, 125).

127 Rosen 1988, 156. Why Rosen calls Gervasoni a "reactionary" theorist may become clear when we consider that Gervasoni's approach contradicts one of the fundamental premises of the modern paradigm of sonata form, namely the dramatic reintroduction of the double return, which is often cited as a criterion for progressiveness.

128 Gervasoni 1800, 467: "[…] e questi ritorni poi servono mirabilmente a rinforzar l'espressione ed a far risovvenire il primo pensiero della Sonata medesima" (transl. in Brover-Lubovsky 2005, 21).

129 Reicha 1826, 298: "[…] le ton de *ré* (le ton principal) et le ton de *la* majeur ne doivent s'y trouver que passagèrement. Le premier, parce qu'il doit prédominer dans la seconde section; le second, parce qu'il a été usé dans la première partie" (transl. in Hoyt 1999, 229). On Reicha's view of sonata form, see also Schmalzriedt 1985.

that Reicha's approach laid the foundation for later descriptions of the development section in which the return to the tonic is more or less forbidden in order not to jeopardize the dramatic moment associated with the recapitulatory beginning.

Taken together, the above statements suggest that the reappearance of the home key shortly after the double repeat sign (or even further into a movement's second half) was not considered to require special explanatory efforts. Despite the large historical distance of about 70 years between such theorists as Riepel and Reicha, eighteenth- and early nineteenth-century music theorists did not view a double return located within the development section as a problematic or deficient event that could only be legitimized by the assumption that the composer wanted to play upon the firmly established convention of the ternary (or Type 3) sonata form. Rather, employing the technique of the medial return was "positively" defined as a cognitive strategy that would provide to the listener a moment of tonal stability from which modulating digressions to more remote tonalities could depart. In addition, composers could thereby help listeners to refresh memory traces of the opening theme. Although the phenomenon of the "omnipresent Meyer" (Brover-Lubovsky[130]) might be difficult to explain from the perspective of modern theories of sonata form, it is entirely in harmony with eighteenth-century conceptions of tonality as 'circular' (or 'solar').[131]

4. Conclusion

The logical consequences to be drawn from the preceding discussion are nothing less than radical: To cast doubt on the very existence of the false recapitulation means to deny the double return the long-range significance typically attributed to it. Since developmental and recapitulatory procedures are much more frequently intermingled than is usually acknowledged, the projective capacity of the double return is consequently assumed to disappear. The double return thus becomes a signpost of purely *local* significance, in a manner similar to a *ritornello*, as Peter Hoyt argues: "If Haydn did not intend to create a deception, then the recurrences of the theme and the tonic in the Classical period were understood as ritornellos and no specific significance concerning formal position or procedure could be reliably attached to them."[132] The characteristic feature of a *ritornello* is that it "communicates nothing that cannot be communicated again."[133] This sets *ritornelli* apart from recapitulatory primary themes, since the latter rarely occur more than once. In addition to its multiple reappearance, a primary theme that is treated as a *ritornello* (irrespective of the tonality in which it appears) may launch a complete rotation of the thematic material found in the exposition, but it need not necessarily do

---

130 Brover-Lubovsky 2005.

131 Ratner 1980, 48.

132 Hoyt 1999, 27. Elsewhere in his dissertation, Hoyt writes: "In analytical discourse, it is rare to speak of more than one recapitulation in a piece" (ibid., 25).

133 Ibid., 25.

so.[134] *Ritornelli* thus do not provide information about the further formal course in the same way that a recapitulatory primary theme would. A *ritornello*-based approach further implies that the main theme need not be restated in full, owing to its repeated use over the course of the movement that has enabled listeners to become familiar with it. It may be presented in a condensed version, according to the synecdochic method described by Hepokoski and Darcy.[135] This *ritornello*-based circular understanding[136] seems to be more appropriate from the historical perspective that views musical forms in the eighteenth century as designed in accordance with various types of interpunction models.[137] Such an understanding runs counter to the modern conception of sonata form as a teleologically unfolding tonal drama culminating in a *singular* moment of (tonal and thematic) return. The feeling of surprise that arises when a double return situated in a sonata form's second half does not reveal itself as the initium of a (complete) recapitulatory rotation and hence "fails" to herald an upcoming moment of resolution may thus ultimately be a surprise without a cause.

References

Adorno, Theodor W. 1976. *Introduction to the Sociology of Music*. Translated by E. B. Ashton. New York: Seabury.

Adrian, Jack. 1990. "The Ternary-Sonata Form." *Journal of Music Theory* 34: 57–80.

——. 1991. "The Function of the Apparent Tonic at the Beginning of Development Sections." *Intégral* 5: 1–53.

Bandur, Markus. 2002. "Plot und Rekurs – 'eine gantz neue besondere Art'? Analytische Überlegungen zum Kopfsatz von Joseph Haydns Streichquartett op. 33, Nr. 1 (Hoboken III:37)." In *Haydns Streichquartette – eine moderne Gattung* (= Musik-Konzepte, vol. 116). Edited by Heinz-Klaus Metzger and Rainer Riehn. Munich: edition text + kritik: 62–84.

Bharucha, Jamshed J. 1987. "Music Cognition and Perceptual Facilitation: A Connectionist Framework." *Music Perception* 5/1: 1–30.

---

134 Galand 2008, 243: "The ritornello as a new point of departure, a potential re-beginning, or, in Hepokoski and Darcy's terminology, the onset of a new rotation – that is the ritornello principle which is fungible here." Most often, main themes are treated in the manner of a ritornello, less often transitional material. I define a "ritornello" as the restatement of any material (P, TR, S, or C) in any tonality (see the TR-ritornello in Mozart's *Prague* Symphony). Galand proposes a narrower definition of the term, using "'ritornello' to identify an opening thematic block and its recurrences, whether or not it might, in certain cases, also be referred to a 'rondo theme' or 'refrain'" (ibid., 242).

135 Hepokoski / Darcy 2006, 233.

136 See Neuwirth 2010. A much more detailed argument in support of a circular model of sonata form can be found in Neuwirth (in preparation).

137 A concise description of the interpunction concept in Reicha's theory can be found in Diergarten 2012, 28–38. The long-standing view of the history of sonata form holds that eighteenth-century theorists tended to privilege tonal procedures, particularly the interpunction order of cadences, over "thematic content" (or "design"), and thus conceived of sonata form as an essentially bipartite structure.

Bonds, Mark Evan. 1988. *Haydn's False Recapitulations and the Perception of Sonata Form in the Eighteenth Century*. Ph. D. diss. Harvard University.

Brover-Lubovsky, Bella. 2005. "*Le diable boiteux*, Omnipresent *Meyer*, and 'Intermediate Tonic' in the Eighteenth-Century Symphony." *Indiana Theory Review* 26: 1–36.

Burstein, Poundie. 1999. "Comedy and Structure in Haydn's Symphonies." In *Schenker Studies* 2. Edited by Carl Schachter and Hedi Siegel. Cambridge: Cambridge University Press: 67–81.

——. 2005. "The Off-Tonic Return in Beethoven's Piano Concerto No. 4 in G Major, Op. 58, and Other Works." *Music Analysis* 24/3: 305–347.

——. 2009. "Echt oder Falsch? Zur tonalen Rolle der 'falschen Reprise' in Haydns Sinfonie Nr. 41." In *Joseph Haydn (1732-1809)*. Edited by Sebastian Urmoneit. Weidler: Berlin: 95–128.

——. 2011a. "True or False? Re-Assessing the Voice-Leading Role of Haydn's So-Called 'False Recapitulations'." *Journal of Schenkerian Studies* 5: 1–37.

——. 2011b. "Schenkerian Analysis and Occam's Razor." *Res Musica* 3: 112–122.

Butt, John. 2010. "Do Musical Works Contain an Implied Listener? Towards a Theory of Musical Listening." *Journal of the Royal Musical Association* 135/1: 5–18.

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.

Clarke, Eric. 1989. "Mind the Gap: Formal Structures and Psychological Processes in Music." *Contemporary Music Review* 3/1: 1–13.

Cone, Edward T. 1968. *Musical Form and Musical Performance*. New York: Norton.

Cook, Nicholas. 1987. "The Perception of Large-Scale Tonal Closure." *Music Perception* 5/2: 197–205.

Cumming, Naomi. 1991. "Analogy in Leonard B. Meyer's Theory of Musical Meaning." In *Metaphor: A Musical Dimension*. Edited by Jamie C. Kassler. Syndey: Currency Press: 177–192.

Czerny, Carl. 1848. *School of Practical Composition*. 3 vols. London: Cocks.

Dahlhaus, Carl. 1987. *Beethoven und seine Zeit*. Laaber: Laaber.

Danuser, Hermann. 1986. "Das imprévu in der Symphonik. Aspekte einer musikalischen Formkategorie in der Zeit von Carl Philipp Emanuel Bach bis Hector Berlioz." *Musiktheorie* 1/1: 61–81.

Diergarten, Felix. 2012. *"Jedem Ohre klingend." Haydns sinfonische Formprinzipien*, Laaber: Laaber.

Dowling, W. Jay / Dane L. Harwood. 1986. *Music Cognition*. New York: Academic Press.

Eerola, Tuomas. 2003. "The Dynamics of Musical Expectancy: Cross-Cultural and Statistical Approaches to Melodic Expectations." PhD dissertation, University of Jyväskylä.

Federhofer, Hellmut. 1981. *Akkord und Stimmführung in den musiktheoretischen Systemen von H. Riemann, E. Kurth und H. Schenker*. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.

Galand, Joel. 2008. "Some Eighteenth-Century Ritornello Scripts and Their Nineteenth-Century Revivals." *Music Theory Spectrum* 30/2: 239–282.

Gervasoni, Carlo. 1800. *La Scuola della Musica*. Piacenza: Niccolò Orcese.

Gjerdingen, Robert. 1999. "An Experimental Music Theory?." In *Rethinking Music*. Edited by Nicholas Cook and Mark Everest. Oxford: Oxford University Press: 161–170.

Grice, H. Paul. 1975. "Logic and Conversation." In *Syntax and Semantics*, vol. 3. Edited by Peter Cole and Jerry L. Morgan. New York: Academic Press: 41–58.

Griesinger, Georg August. 1810. *Biographische Notizen über Joseph Haydn*. Leipzig: Breitkopf & Härtel.

Haimo, Ethan T. 1995. *Haydn's Symphonic Forms: Essays in Compositional Logic*. Oxford: Oxford University Press.

Hepokoski, James. 2002. "Beyond the Sonata Principle." *Journal of the American Musicological Society* 55/1: 91–154.

Hepokoski, James / Warren Darcy. 1997. "The Medial Caesura and Its Role in the Eighteenth-Century Sonata Exposition." *Music Theory Spectrum* 19/2: 115–154.

——. 2006. *Elements of Sonata Theory*: *Norms, Types, and Deformations in the Late-Eighteenth-Century Sonata*. New York: Oxford University Press.

Hinrichsen, Hans-Joachim. 2004. "Durchkreuzte und eingelöste Erwartung. Schuberts Kammermusik." In *Dem Ohr voraus. Erwartung und Vorurteil in der Musik* (= Studien zur Wertungsforschung 44). Edited by Andreas Dorschel. Vienna: Universal Edition: 112–133.

Hippel, Paul von / David Huron. 2000. "Why do Skips Precede Reversals? The Effect of Tessitura on Melodic Structure." *Music Perception* 18/1: 59–85.

Hoyt, Peter A. 1999. *The 'False Recapitulation' and the Conventions of Sonata Form*. Ph. D. Diss. University of Pennsylvania.

Huron, David. 2006. *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge/Mass: MIT Press.

Jones, Mari Riess. 1987. "Dynamic Pattern Structure in Music: Recent Theory and Research." *Perception and Psychophysics* 41: 621–634.

Kim, Jin-Ah. 2004. "Was vermag die Sonatentheorie heute noch zu leisten? Zur Rehabilitierung eines fragwürdig gewordenen Formbegriffs." In *Musiktheorie zwischen Historie und Systematik. Bericht über den 1. Kongreß der Deutschen Gesellschaft für Musiktheorie, Dresden 2001*. Edited by Ludwig Holtmeier, Michael Polth, and Felix Diergarten. Augsburg: Wissner: 121–130.

Kirnberger, Johann Philipp. 1771. *Die Kunst des reinen Satzes in der Musik*. Reprint Kassel: Bärenreiter 2004 [English translation by David Beach and Jurgen Thym. 1982. *The Art of Strict Musical Composition by Johann Philipp Kirnberger. Introduction and Explanatory Notes by David Beach*. New Haven: Yale University Press].

Korvstedt, Benjamin. 2013. "On Not Inviting Difficulties in Haydn's Symphonies." *HAYDN: Online Journal of the Haydn Society of North America* 3/2. http://haydn-journal.org

Lakoff, George. 2008. *Women, Fire, and Dangerous Things*. Chicago: University of Chicago Press.

La Motte-Haber, Helga de. 1995. "Der einkomponierte Hörer." In *Der Hörer als Interpret*. Edited by Helga de la Motte-Haber / Reinhard Kopiez. Frankfurt a. M.: Lang: 35–41.

Landon, H. C. Robbins. 1955. *The Symphonies of Joseph Haydn*. London: Universal Edition.

Larsen, Jens Peter. 1963. "Sonatenform-Probleme." In *Festschrift Friedrich Blume zum 70. Geburtstag*. Edited by Anna Amalie Abert / Wilhelm Pfannkuch. Kassel: Bärenreiter: 221–231.

Larson, Steve. 2002. "Musical Forces, Melodic Expectation, and Jazz Melody." *Music Perception* 19/3: 351–385.

——. 2003. "Recapitulation Recomposition in the Sonata-Form First Movements of Haydn's String Quartets: Style Change and Compositional Technique." *Music Analysis* 22/1–2: 139–177.

LaRue, Jan. 1992. "Bifocal Tonality in Haydn's Symphonies." In *Convention in Eighteenth- and Nineteenth-Century Music: Essays in Honor of Leonard G. Ratner*. Edited by Wye J. Allanbrook, Janet m. Levy, and William P. Mahrt. Stuyvesant, New York: Pendragon Press: 59–73.

Leichtentritt, Hugo. 1927. *Musikalische Formenlehre*. Third Edition. Leipzig: Breitkopf [English translation by Hugo Leichtentritt. 1951. *Musical Form*. Cambridge, Mass.: Harvard University Press].

Levy, Janet. 1982. "Texture as a Sign in Classic and Early Romantic Music." *Journal of the American Musicological Society* 35/3: 482–531.

Lewin, David. 1986. "Music Theory, Phenomenology, and Modes of Perception." *Music Perception* 3/4: 327–392.

Löhlein, Georg Simon. 1765/1782. *Clavier-Schule, oder kurze und gründliche Anweisung zur Melodie und Harmonie, durchgehends mit practischen Beyspielen erkläret*. Leipzig and Züllichau: Frommann.

London, Justin. 1996. "Musical and Linguistic Speech Acts." *The Journal of Aesthetics and Art Criticism* 54/1: 49–64.

Margulis, Elizabeth Hellmuth. 2005. "A Model of Melodic Expectation." *Music Perception* 22/4: 663–714.

Marvin, Elizabeth West / Alexander Brinkman. 1999. "The Effect of Modulation and Formal Manipulation on Perception of Tonic Closure by Expert Listeners." *Music Perception* 16/4: 389–407.

Maus, Fred. 1999. "'Music in the Moment': A Discussion." *Music Perception* 16/4: 463–494.

Meyer, Leonard B. 1956. *Emotion and Meaning in Music*. Chicago: Chicago University Press.

——. 1967. *Music, the Arts, and Ideas: Patterns and Predictions in Twentieth-Century Culture*. Chicago: Chicago University Press.

——. 1973. *Explaining Music: Essays and Explorations*. Berkeley: University of California Press.

——. 1989. *Style and Music: Theory, History, and Ideology*. Philadelphia: University of Chicago Press.

Narmour, Eugene. 1977. *Beyond Schenkerism*: *The Need for Alternatives in Music Analysis*. Chicago: University of Chicago Press.

——. 1990. *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. Chicago: University of Chicago Press.

——. 1992. *The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model*. Chicago: University of Chicago Press.

——. 1999. "Hierarchical Expectation and Musical Style." In *Psychology of Music*. Edited by Diana Deutsch. New York: Gulf Professional Publishing: 441–472.

Neuwirth, Markus. 2009. "Der mediantische Reprisenübergang bei Joseph Haydn und einigen seiner Zeitgenossen zwischen Konvention und Normverstoß: Satztechnische Inszenierung, formale Implikationen und Erklärungsmodelle." *ZGMTH* 6/2: 231–271.

——. 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: 369–385.

——. 2011. "Joseph Haydn's 'Witty' Play on Hepokoski and Darcy's *Elements of Sonata Theory*." *ZGMTH* 8/1: 199–220.

——. in preparation. *"New Twists of the Old": Recomposed Recapitulations in the Works of Joseph Haydn and His Contemporaries*. Leuven: Leuven University Press.

Ratner, Leonard G. 1949. "Harmonic Aspects of Classic Form." *Journal of the American Musicological Society* 2/3: 159–168.

——. 1980. *Classic Music: Expression, Form, and Style*. New York: Schirmer Books.

Ratz, Erwin. 1968. *Einführung in die musikalische Formenlehre: Über Formprinzipien in den Inventionen und Fugen J. S. Bachs und ihre Bedeutung für die Kompositionstechnik Beethovens*. Wien: Universal Edition.

Reicha, Antonin. 1824–1826. *Traité de haute composition musicale*. Paris: Zetter.

Riepel, Joseph. 1755. *Anfangsgründe zur musicalischen Setzkunst* (= Sämtliche Schriften zur Musiktheorie). Edited by Thomas Emmerich. 2 Vol. Vienna: Reprint Böhlau 1996.

Rosen, Charles. 1988. *Sonata Forms*[2]. New York: Norton.

——. 1997. *The Classical Style: Haydn, Mozart, Beethoven*. New York: Norton.

Rosen, David. 1996. "'Unexpectedness' and 'Inevitability' in Mozart's Piano Concertos." In *Mozart's Piano Concertos: Text, Context, Interpretation*. Edited by Neal Zaslaw. Ann Arbor: The University of Michigan Press: 261–284.

Rywosch, Bernhard. 1937. "Der Übergang zur Reprise in Haydns und Mozarts früherer Symphonik." *Schweizerische musikpädagogische Blätter* 26: 115–118 and 132–135.

Schellenberg, E. Glenn. 1997. "Simplifying the Implication-Realization Model of Melodic Expectancy." *Music Perception* 14/3: 295–318.

Schenker, Heinrich. 1954. *Harmony*. Edited by Oswald Jonas, translated by Elizabeth Mann Borgese. Chicago: University of Chicago Press. Translation of *Harmonielehre*. Stuttgart: Cotta, 1906.

——. 1979. *Free Composition*. Edited and translated by Ernst Oster. New York: Longman.

Schmalzriedt, Hugo. 1985. "Charakter und Drama: Zur historischen Analyse von Haydnschen und Beethovenschen Sätzen." *Archiv für Musikwissenschaft* 42/1: 37–66.

Schmidt-Thieme, Lars. 2000. *Die formale Gestaltung von Exposition und Reprise in den Streichquartetten Haydns*. Frankfurt a. M.: Peter Lang.

Schmuckler, Mark. 1989. "Expectation and Music: Investigation of Melodic and Harmonic Processes." *Music Perception* 7/2: 109–149.

Schoenberg, Arnold. 1967. *Fundamentals of Musical Composition*. Edited by Gerald Strang / Leonard Stein. London-Boston: Faber & Faber.

Sisman, Elaine. 1993. *Mozart: The "Jupiter" Symphony*. Cambridge: Cambridge University Press.

Somfai, László. 1995. *The Keyboard Sonatas of Joseph Haydn: Instruments and Performance Practice, Genres and Styles*. Chicago: University of Chicago.

Spitzer, Michael. 1996. "The Retransition as Sign: Listener-Oriented Approaches to Tonal Closure in Haydn's Sonata-Form Movements." *Journal of the Royal Musical Association* 121/1: 11–45.

Stöhr, Richard. 1927. *Musikalische Formenlehre*. Third edition. Leipzig: Siegel.

Strunk, W. Oliver. 1932. "Haydn's Divertimenti for Baryton, Viola, and Bass (After Manuscripts in the Library of Congress)." *The Musical Quarterly* 18/2: 216–251.

Thomson, William. 1983. "Functional Ambiguity in Musical Structures." *Music Perception* 1/1: 3–27.

Tillmann, Barbara / Emmanuel Bigand / Francois Madurell. 1998. "Local versus Global Processing of Harmonic Cadences in the Solution of Musical Puzzles." *Psychological Research* 61: 157–174.

Tobel, Rudolf von. 1935. *Die Formenwelt der klassischen Instrumentalmusik* (= Berner Veröffentlichungen zur Musikforschung IV). Bern/Leipzig: Haupt.

Tovey, Donald Francis. 1935. *Essays in Musical Analysis*, vol. 1: *Symphonies*. London: Oxford University Press.

——. 1944. "Sonata Forms." In *Musical Articles from the Encyclopaedia Britannica*. Edited by Hubert J. Foss. London: Oxford University Press: 208–232.

Webster, James. 1986. "Binary Variants of Sonata Form in Early Haydn Instrumental Music." In *Internationaler Joseph Haydn Kongress Wien 1982*. Edited by Eva Badura-Skoda. Munich: Henle: 127–135.

——. 2001. Art. "Sonata Form." In *The New Grove Dictionary of Music and Musicians* 17. Edited by Stanley Sadie and John Tyrell: 687–701.