If cyclic composition is the greatest musical achievement of the late 18th century — and the symphonic and chamber music of Haydn, Mozart and Beethoven continues to occupy a pre-eminent place in Western music culture, despite recent trends towards destabilising the canon — then the study of sonata form is perhaps the highest goal one can pursue in music theory. And a wide range of musicians — performers and composers, as well as theorists, historians and journalists — continue to aspire to this goal, whether they do so intuitively or in a systematic fashion.

Pursued systematically, sonata theory mediates between the particular and the general, between the individual qualities of a given work and the principles derived from a large body of music. A theory must be able not merely to identify what is individual but to celebrate what it is that makes that individuality worth cherishing; to do this requires a set of concepts — variously called ‘laws’, ‘rules’, ‘regulative principles’, even ‘axioms’ — by which to make technical observations and aesthetic judgements. Sonata pieces contain elements of originality as well as exhibiting features of musical coherence; a sonata theory — a good one — must be able to account for both.

The term ‘sonata form’ was first coined in the first half of the 19th century: Adolf Bernhard Marx used it as early as 1823, the first year of publication of his Berliner allgemeine musikalische Zeitung, and more formally from 1838, in Die Lehre von der musikalischen Komposition. Since then, but more particularly from about 1900, scholars have recognised the difficulty of devising a theory that accounts for the fact that many pieces containing original or novel elements will eventually be accepted into the canonic repertory of musical works and that, as a result, elements previously classified as lying outside or contrary to the theory will gain acceptance as part of an enlarged repertory of compositional technique. Thus the concept of musical coherence, in order to keep up with the ever-increasing totality of the repertory, is constantly in flux.

I shall give one famous — overly famous — example of this later on, but let us first look at the basics of sonata form according to James Hepokoski & Warren Darcy as elaborated in their new book.

For Hepokoski and Darcy, sonata movements are trajectories of musical energy across a series of ‘zones’ (also called ‘action-zones’ or ‘action-spaces’), in which listeners may reasonably expect certain things to happen. What we — and also the authors — call the exposition is divided into four zones: a
'primary theme zone' (P), which establishes the tonic key; a 'transition' (TR), described as an 'energy-gaining zone', which leads to a new ('non-tonic') key; a 'secondary theme zone' (S) in the new key; and, usually but not obligatorily, a 'closing zone' (or 'closing space') (C).

These zones are controlled by trajectories operating at various levels. The most far-reaching of these extends from the beginning of the primary theme zone in the exposition to the end of the secondary theme zone in the recapitulation, i.e. the point at which the tonic key is confirmed by a strong perfect authentic cadence in that key (represented as I:PAC); this point is designated the 'essential structural closure' (ESC). There are also important trajectories at work within the exposition and the recapitulation, leading from the start of the primary to the end of the secondary theme zone in each of these sections. By analogy with the ending, Hepokoski & Darcy posit an 'essential expositional closure' (EEC) at a comparable point in the exposition, where there is a perfect authentic cadence at the end of the secondary theme zone, normally in the key of the dominant (V:PAC). Finally, there are lower-level trajectories across the two secondary theme zones, leading again to the essential expositional close and the essential structural close, respectively. The basics of the scheme, the 'generic layout of sonata form', are given in their fig.2.1, reproduced here as fig.1.

An important constituent of the authors' model is the break in the exposition, and also in the recapitulation, before the arrival of the secondary theme zone. This they term the 'medial caesura' (MC); it is a feature of sonata form that helps the listener understand the secondary theme as a new beginning—a 'relaunch', as they put it—often in a more lyrical vein in comparison with the primary theme. (This term has become a Hepokoski & Darcy trademark: the medial caesura was the subject of a jointly-authored article in *Music Theory Spectrum* a decade ago, which spawned the present project.)

As can be seen from the pair of diagrams in fig.1, the exposition and recapitulation are not only elaborate schemes in themselves, they form a kind of partnership, one that is expressed here in terms of 'promise' and 'accomplishment'. Accomplishment is, simply, closure, the return of the music to its harmonic and thematic origins; what makes the voyage exciting is the distance covered by the music during this process. By contrast, the action-zone that remains known as the development has the fewest prescriptions attached to it: its attributes are suggested, not obligatory. As can be seen from fig.1b, it is 'often' dominated by thematic material from the primary zone or transition or 'perhaps rotational', by which the authors mean that it may be based on a succession of several themes from the exposition. The diagram does not actually tell us what the development is supposed to achieve, other than to separate the exposition from the recapitulation; but that is, simply, all that it needs to do: by putting distance (i.e. time) between the two outer
sections, it increases the sense of excitement in the listener who expects the music from the beginning of the piece to make its return.

There are, of course, many variants and variations to the generic layout. After a lengthy and detailed account of the action-zones — at least one substantial chapter is devoted to each — and substantial chapters on the development and recapitulation, Hepokoski & Darcy move on to related issues: (1) codas and introductions: these are designated as ‘parageneric spaces’, because they are not essential components of the ‘genre’ of sonata form but were used with increasing frequency in the 18th century ‘as add-ons to the basic structure’; (2) sonata forms whose tonics are minor keys, which they treat as a special case of sonata form; (3) sonatas as cycles of more than one movement, i.e. sonata form as part of a larger entity; (4) forms related to the standard, ‘textbook’ sonata form: sonata without development (sometimes known as ‘sonatina form’), binary forms, sonata-rondos, and concerto first-movement form.
Mostly Mozart

The authors round off their exposé with two supplementary chapters, designated as appendices. The first, ‘Some grounding principles of sonata theory’, is in effect an epistemology of sonata form, showing the relationship between the theory of late 18th-century music and the cultural and social conditions leading to their multi-faceted production. The second glosses two key concepts of Hepokoski & Darcy’s theory: ‘rotation’, which concerns sonata theory as a theory of music based on ‘cycles’ of thematic material; and ‘deformation’, the process by which the universe of sonata pieces is enriched by the composition of works that enlarge the scope of a listener’s expectation.

In dealing with the multitude of variety among sonata pieces from the Classical period, the authors adapt their basic vocabulary to explain particular situations as they arise; or a new term may be introduced if the original model is not sufficiently flexible. As is well known, an action-zone may be based on a multitude of ideas, rather than on a single ‘subject’, and these should be identified accordingly. Thus, for example, a series of ideas in the secondary key could be labelled $S'$, $S''$, $S'''$. If there are grounds for associating some of these more closely with each other, superscripts should be appropriately modified, e.g. $S'$, $S''$, $S'''$. Sometimes the first secondary theme is introduced by a ‘preparatory module’ lacking in thematic profile, which may be designated as $S''$, or even $S'''$; or the last of the group might have the character of a closing theme, which could be represented as $S''$.

There are repertories – the late 18th-century concerto, and the early Romantic sonata, for example – for which it is preferable to posit a second medial caesura, resulting in a more elaborate form of action-zone for the themes in nontonic keys: a new theme after the first caesura, a process of development towards the second, and a further new theme. Terming the phenomenon the ‘apparent double medial caesura’ to emphasise the unity of the combined effect of the two breaks, the authors define the entirety of the music thus set up as a ‘trimodular block’ (TMB) and label the components thereof as $TM'$, $TM''$, $TM'''$, or with a more sophisticated set of superscripts if, say, there are more themes, or thematic interrelationships (e.g. $TM^{4}$, $TM^{5}$). Concertos have further special features, including a series of ‘ritornellos’ (R) and ‘solo sections’, labelled $S'$, $S''$ and $S'''$ to avoid confusion with the letter $S$ as used for the secondary theme or action-zone.

If the above explanations seem heavily laden with terms and abbreviations, that is merely to give a taste of Hepokoski & Darcy’s style. Their text requires the reader to absorb and assimilate terms as soon as they are introduced, since the later elaborations of the theory depend on them. As a mild example of this language, I quote the beginning of the authors’ discussion of the transition section (TR) as it reappears in a sonata-form recapitulation. (In what follows, HC stands for ‘half cadence’ – in British parlance, ‘imperfect cadence’.)
The Recapitulatory TR

It was common for a composer to modify the recapitulatory P and/or TR zones. The TR, in particular, was a passage that invited recomposition. Alterations were all the more apt if the exposition’s TR had been a modulatory transition, that is, if it had driven toward a V:HC MC (or III:HC MC or vi:HC MC in a minor-mode sonata). That same modulatory move was not to be made in the recapitulation’s TR. Since S was now to emerge in the tonic, the MC endpoint of TR would have to be a I:HC instead, down a fifth from that of the exposition.

All that this says is that, if a transitional passage had set up the second subject by moving to the dominant of the secondary key, then in the recapitulation it must move to the dominant of the home key. But wherever a technical term can be used, the authors will use it; and if a term can be abbreviated, they will use its abbreviation. Readers will find the text slow-going, and impossible either to skim or drop at one point and pick up later on. An alphabetical list of terms and abbreviations — in effect, a glossary for the theory — precedes the main text, on pp. xxv–xxviii, introducing readers to the elaborate system of terminology before it has been explained.

The exciting, challenging, part of sonata theory results from the tension between general principles, which can be derived from the repertory as a whole, and the exceptions one is forever encountering in individual works. Here the authors adopt a number of related strategies. The introduction of the concept of deformation is preceded by that of the more benign, gentler notion of ‘default’ — the scenario one is most likely to encounter — so named by analogy with the default settings on a computer or similar piece of equipment. A default setting not only is the most probable outcome, it implies the existence of an alternative, possibly further alternatives, down the line (hence the notion of ‘primary’ and ‘secondary’ default settings). The notion of default is especially applicable to the medial caesura, as there are a number of paths from the tonic to the contrasting key that composers familiar with the music of their time could choose from. But sometimes all the alternatives are rejected, or there simply are no alternatives yet known, in which case a composer may effect a ‘deformation of that compositional moment’: a departure from all known options, the choice of a path never before taken. But as soon as a normative procedure is subjected to a deformation (the verb ‘to deform’ and the adjective ‘deformed’ are studiously avoided), the new procedure has the potential to become a secondary default.

Numerous examples by Beethoven come readily to mind, especially in the period 1800–04, a time during which many time-honoured sonata features were abandoned, including the dominant as key for the secondary action-zone. But just how clearly is the process of deformation — understood as a rejection of the normative — at work in this repertory? Here is Hepokoski & Darcy in historical mode, writing on the subject of modulation to the mediant major:
Around 1800 Beethoven began to investigate the deformation of moving to III for part 2, not to V, within major-mode sonata-based structures. This might have been suggested by analogy to the role of the mediant in minor-mode sonatas. Examples may be found in the first movements of his Piano Sonatas in G, op.31 no.1 (S first in B major, m.66, then repeated in B minor, m.74, which key persists for the rest of the exposition) and in C, op.53 ('Waldstein,' S in E major), in the Leonore Overtures Nos.2 and 3 (P in C major; S and C in E major), and in the finale of the Piano Trio in E-flat, op.70 no.2 (S and C in G major).

This account is factually accurate; and the authors then go on to discuss an alternative tonal route, from I to VI, that Beethoven first tried out in some of his lesser known instrumental works from the period. But there remains the question ‘why?’: why III, rather than V, as the goal of the first transition in the ‘Waldstein’ Sonata? What would have been wrong with a ‘Waldstein’, other than that the Beethoven of 1803–04 wrote otherwise, whose first theme took the path shown in ex.1 when it was restated in bars 14ff.

Is there something about the chromatic bass line C–C♯–D that is technically inept? If not, was Beethoven either being wilful by taking a different route, or there is something about the first subject that directs the restatement towards the key of E? Is it perhaps the B♭ in bar 5 that ‘motivates’ an A♯ in bar 22 (in Beethoven’s version), thus turning an intrinsically flatward-sinking theme towards the sharp side of C major with greater persuasiveness?

Hepokoski & Darcy do not tell us why the ‘Waldstein’ goes to E major — Elements of sonata theory does not like to answer this sort of question — but have instead offered another explanation of its tonal plan: the establishment of a secondary default by an earlier work, the Sonata in G op.31 no.1 (1802), as shown in ex.2. The F♯ major chord straddling bars 53–54 effects a deformation in the transition section. And by moving to B major, rather than to the predicted key of D in a convincing way, Beethoven has created a new, secondary, default setting, which he will choose for the ‘Waldstein’ the following year.

And the story continues. At the time of his late sonatas and quartets, it is just as likely for a Beethoven sonata exposition to modulate to a third-related tonality — e.g. A major (III) in the finale of the Quartet in F op.135 — as it is to the dominant. By the 1820s, V is no longer the primary default setting for the second subject.

Even if that is the case, one still wants to know why Beethoven took the ‘wrong’ turning in op.31 no.1. Does the element of subversion have a greater role to play in a sonata whose main idea is full of stops and starts, of surprises? And yet, even if this were a possible explanation of the secondary action-zone in the key of B, why did Beethoven fail to take the same route a few years before when, in a similarly mercurial, witty sonata, op.10 no.2, in F, he had opened up a path that would have led just as convincingly to III (ex.3)? Why, after arriving at this point, did he plunge into C major, instead of continuing in the key of A? Might it be equally plausible to view the
Ex.1: Beethoven: Piano Sonata in C op.53 ("Waldstein"), first movement, bars 14–42, recomposed

Allegro con brio
exposition of op. 10 no. 2 as undergoing ‘deformation’ by rejecting the seemingly inevitable path to the mediant and, instead, moving to the conditionally less likely (if historically more probable) key of the dominant?

Examples of this sort lead me to wonder whether Hepokoski & Darcy are entirely comfortable as music historians, alongside their role as theorists. Their text is, as I have shown, dominated by new vocabulary and technical terminology; and while the bibliography bears eloquent witness to the range and depth of their reading, they seem more reluctant to acknowledge the significance of 20th-century scholarship than to argue its shortcomings, or limited applicability. Take, for instance, the notion of ‘sonata principle’ first articulated by Edward Cone in 1968, according to which important thematic material first presented outside the home key should be recapitulated in the tonic, and Charles Rosen’s closely allied notion of ‘structural dissonance’ (1972), the large-scale harmonic conflict created whenever thematic material is stated in a secondary key, i.e. the notion that thematic material, though not intrinsically dissonant, nevertheless calls for some higher-level ‘resolution’. By observing that these concepts do not quite fit a number of late 18th-century works, Hepokoski & Darcy seem merely to be reserving the deformation game for their own purposes: it is, after all, crucial to an understanding of the ‘Eroica’ first movement that the ‘new theme’ in the development, in the distant key of E minor, is restated in the tonic minor and, similarly, that the chorale-like second subject of the ‘Waldstein’ Sonata is heard in the home key, C major, towards the end of the coda. Nor are Hepokoski & Darcy right to dismiss as an ‘eccentric claim’ James Webster’s important suggestion that the ‘double return’ (explained in his ‘Sonata form’ entry in the New Grove) — the coincidence of the first theme and the return of the home key at the point of recapitulation — was a crucial feature in the evolution of sonata form from earlier binary designs, even if it is sometimes missing from pieces we shall always want to call sonatas, e.g. the first movement of Mozart’s famous Sonate facile in C major K. 545.

Wherever they come across some music-theoretical concept that is either in conflict with one of their own or too closely related to ignore, they manage to find grounds for qualifying its earlier usage or rejecting it outright. Thus, for example, Robert Winter’s term ‘bifocal close’ — the unmediated reinterpretation of the dominant chord as new key — is dismissed, without explanation, as simply ‘not [...] helpful’. And William Caplin’s ‘expanded cadential progression’ is dealt with summarily in a pair of footnotes. By contrast, they seem willing to adopt Ralph Kirkpatrick’s notion of ‘crux’, from his monograph on Domenico Scarlatti (1953), provided that the term is applied only to the recapitulation, i.e. to mark the point at which the music of a sonata exposition moves irrevocably away from the path it had taken previously as it sought to reach the contrasting key. Of course, no single writer — Marx,
Schenker, Cone, Rosen, Webster, Caplin – has come up with an explanation of sonata form that is entirely satisfactory, but there is general agreement that our understanding of the theory has evolved over time; Hepokoski & Darcy seem, however, reluctant to cast themselves in a role in that evolutionary process.

As for historical sources of sonata form, the theoretical literature of the late 18th and early 19th centuries is discussed wherever it is consistent with the theories propounded here, but the authors insist that the writings of 'Koch, Galeazzi, Reicha, and so on' must ultimately be set aside in favour of the 'more robust quality of information' gained by a study of 'what the masters do in real composition'. (Beethoven makes a guest appearance as a theorist, but his sketchbook annotations tell us much more about his con-
Mostly Mozart

Mostly Mozart

The reliance upon music, rather than on writings about music, is understandable enough, but it does not square with the claim made in the Preface that this book is ‘a blend of musicological and music-theoretical thinking’. (And the authors are not entirely clear about what constitutes ‘real composition’ by ‘masters’: works by Boccherini, Cherubini, Dittersdorf, Vanhal, Wagenseil and Weber are mentioned, but none is illustrated.)

Of course, certain strands of theoretical thinking are too elusive, or sacred, for the authors to risk hammering. Donald Tovey, one of the most important writers on cyclical forms (including concerto form), escapes censure, perhaps because his theories were cleverly disguised as concert programme notes. So does Heinrich Schenker, though many card-carrying Schenkerians believe that his concepts of musical form were expressed inadequately in the New musical theories and fantasies (a projected volume on form was never written). Even here, Hepokoski & Darcy restrict themselves to the final formulation of Schenkerian doctrine, the posthumously published Der freie Satz (1935), whose section on musical form appears to have been hastily drawn up. They have chosen not to look at his 1923 trilogy of essays on piano sonatas by Haydn, Mozart and Beethoven, or at any of his monumental analyses of symphonies by Mozart (G minor) and Beethoven (nos. 3, 5 and 9), all of which were available not only in the original German but also in reliable English translations at the time Elements of sonata theory was in preparation. (Astonishingly, not even Schenker’s 1926 essay ‘On organicism in sonata form’, twice translated into English, is listed in the bibliography.)

Given the enormous scope of the book, one would expect the examples to be drawn from a wide range of works. But while Hepokoski & Darcy’s knowledge of the repertory is impressively indexed, actual musical illustrations are restricted almost entirely to the works of Haydn, Mozart and Beethoven: there is one example each from sonatas by Domenico Scarlatti, Johann Christian Bach and Carl Philipp Emanuel Bach. And within the group of Austrian masters, the distribution is uneven: Haydn 7, Mozart 56, Beethoven 11.

This observation is intended neither to cast doubt upon Mozart’s mastery of sonata-form procedures – no other composer of the period wrote so imaginatively across so wide a range of genres – or to imply that the authors are more familiar with his work than with that of his contemporaries; but since they devote so much space to Mozart’s piano concertos in the final chapters of the book, they ought have shifted the emphasis in their discussion of ‘textbook’ sonata form towards the work of other composers: and even if Haydn’s inventiveness as a sonata composer is duly acknowledged, the short-
age of concrete illustrations undermines his position as the 18th-century deformation supremo. Moreover, the book gives the impression that the authors are not fully up-to-date with Haydn scholarship. In a discussion of the influence of musical topoi on form, their reference to 'Haydn's nick-named symphonies' implies – quite erroneously – that these nicknames are largely authentic. And in explaining Haydn’s ingeniously understated recapitulation in the Quartet in B♭ op.50 no.1, they are so distracted by the opportunity to score points against a triumvirate of leading Haydn scholars—Janet Levy, Dean Sutcliffe and Charles Rosen — that they fail to take into account the relevant textual scholarship and blithely quote the extract as it appears in the early 20th-century Eulenburg and Peters editions.

The tone of the book is, to my ears, a shade too confrontational for what has been conceived as a work of reference. The appropriation of terminology (e.g. ritornello ‘pillars’ from James Webster, ‘crux’ from Kirkpatrick, even ‘sujet libre’ from the early 20th-century French team of Théodore de Wyzewa & Georges de Saint-Foix) is acknowledged, yet one gets the impression that this has been done with some reluctance, that Hepokoski & Darcy are — groundlessly — irritated that someone got there first. Elsewhere, they are willing almost to invent scholarly disputes, e.g. over the term ‘phrase’ as used by William Caplin and William Rothstein and, later, over the same two authors’ understanding of essential expositional/structural closure and the passage leading up to it; some battlefield metaphors (‘to shore up his own position’, ‘had stood on firmer historical ground’) seem out of place here. There is also a gratuitous fuss over a misprint in, of all places, a volume from the Neue Mozart-Ausgabe.

The bibliography is comprehensive: the only major omission I spotted, apart from Schenker’s sonata-form essay and analyses, was Charles Smith’s magisterial study of Schenker’s concept of form, published in Music Analysis in 1996. It is followed by three indexes, of which the second (works cited) is the most straightforward. (Again we see the predominance of Mozart, with over ten columns, as against eight for Haydn and Beethoven between them.)

The index of names is detailed, but the association of terms and concepts with scholars is inconsistent. Under Webster one finds ‘double return’; but ‘sujet libre’ is found neither under Wyzewa nor under Saint-Foix; ‘sonata principle’ is not found under Cone; and under Rosen, ‘dissonance of expositional nontonic material’ appears in place of the simpler, more familiar ‘structural dissonance’. (I could find nothing in the larger Index of Concepts that corresponds to this expression.) There are a generous number of subheadings under Caplin, but none for his seminal ‘expanded cadential progression’ (in the Index of Concepts, this can be found under ‘cadence’).

Although the final, and longest, Index of Concepts comes with a note warning us not to expect every occurrence of every term to be referenced, it
Mostly Mozart gives us a good idea of the immense scale of Hepokoski & Darcy’s undertaking, and the problems engendered by their elaborate system of terms and abbreviations. The use of headings governing a very long list of subheadings often makes the task of finding things difficult, especially at the letters R, S, and T in the alphabet. Under ‘concerto’, a topic that takes up a large part of the book, readers are sent to ‘Type 5 sonata’, at which point they must choose from the following main headings (not subheadings), which appear logically, if not intuitively, in the alphabetical order: Type 5 sonata (concerto); Type 5 sonata, Ritornello 1 (R1); Type 5 sonata, Ritornello 2 (R2); Type 5 sonata, Ritornello 3 (R3); Type 5 sonata, Ritornello 4 (R4); Type 5 sonata, Solo 1 (S1); Type 5 sonata, Solo 2 (S2, developmental space); Type 5 sonata, Solo 3 (S3, solo recapitulation). The complexity and unwieldiness of the Index of Concepts may, however, have the unintended benefit of putting readers into browsing mode, where their curiosity will be rewarded by the appearance of such intriguing entries as ‘bait-and-switch tactic’, ‘“lights-out” effect’, ‘prison-house effect’ and others whose meanings are far from self-evident.

This is a long book: I estimate that there are some 300,000 words in its 22 double-columned chapters and two supplementary appendices on terminology. As difficult as it may be to ignore what the authors have achieved here, their text cannot easily be digested in a short time-span, nor is it one that I can imagine easily put down and taken up at will. While the last four chapters, on Mozart’s concerto first movements, virtually constitute a study in its own right, it cannot be read independently of what precedes it: the technical terminology from the first part of the book is carried over into the second, and there are many references to concertos in the chapters on ‘textbook’ sonata form.

Its technical language, its uneven and occasionally discordant tone, its bias toward Mozart, and above all its sheer size notwithstanding, this challenging, wide-ranging study is still worth tackling. In spite of the reservations voiced here, the authors’ achievements are impressive, even daunting. It is too early to predict whether the Elements of sonata theory will become an authoritative statement on sonata form, to which future generations of students will refer. That may well depend on whether today’s generation is willing to make the sustained effort of coming to terms with it.

William Drabkin is Professor of Music at the University of Southampton. His publications include English editions of Schenker’s Das Meisterwerk in der Musik and Der Tonwille.
Copyright of Musical Times is the property of Musical Times Publications Ltd. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.