Popular Music (2011) Volume 30/1. Copyright © Cambridge University Press 2011, pp. 47–70
doi:10.1017/S026114301000067X

A corpus analysis of rock harmony

TREVOR DE CLERCQ and DAVID TEMPERLEY

Eastman School of Music, Department of Music Theory, 26 Gibbs Street, Rochester, NY 14604, USA
E-mail: trevor.declercq@rochester.edu
E-mail: dtemperley@esm.rochester.edu

Abstract

In this study, we report a corpus analysis of rock harmony. As a corpus, we used Rolling Stone *magazine's list of the '500 Greatest Songs of All Time'; we took the 20 top-ranked songs from each decade (the 1950s through the 1990s), creating a set of 100 songs. Both authors analysed all 100 songs by hand, using conventional Roman numeral symbols. Agreement between the two sets of analyses was over 90 per cent. The analyses were encoded using a recursive notation, similar to a context-free grammar, allowing repeating sections to be encoded succinctly. The aggregate data was then subjected to a variety of statistical analyses. We examined the frequency of different chords and chord transitions. The results showed that IV is the most common chord after I and is especially common preceding the tonic. Other results concern the frequency of different root motions, patterns of co-occurrence between chords, and changes in harmonic practice across time.*

Introduction

In the extensive scholarship on rock that has emerged over the last two decades, harmony has been a central concern. Many analytical studies of individual songs have focused primarily or even exclusively on harmony and other aspects of pitch organisation (see for example many of the essays in Covach & Boone 1997, Holm-Hudson 2002 and Everett 2008a). More general studies of the style, too, have devoted much of their attention to harmony (a few examples will be discussed below). Indeed, some have argued that rock scholarship has focused excessively on harmony and pitch organisation at the expense of other, less tractable musical dimensions such as timbre and micro-level nuances of pitch and rhythm (Tagg 1982, pp. 41–2; Middleton 1990, pp. 104–5). While we concede this point, it is surely uncontroversial to state that harmony is *one* important aspect of rock, and that an understanding of this aspect is necessary for a full understanding of the style as a whole.

In any harmonic style, the sequence of chords in a piece is presumably not just a series of random selections from the entire vocabulary (whatever that may be – the set of all major and minor triads, for example); rather, certain chords are preferred over others, and certain patterns of motion (chord-to-chord progressions and larger patterns) are also favoured. A well known case in point is common-practice music (Western art music of the 18th and 19th centuries). The conventional wisdom of tonal music theory offers us a set of basic principles about common-practice harmony – not, of course, inviolable and exceptionless laws, but general norms that apply most of the time. In their essential form, these principles can be stated quite succinctly. Harmonies are grouped into three categories of tonic (I), pre-dominant (IV and II) and dominant (V and VII) (the status of VI and III in this scheme is dependent on context); generally speaking, pre-dominants move to dominants, dominants move to tonics, and tonics can move to any category. Another widely accepted principle is that certain root motions are particularly favoured, especially ascending motion by fourths.[1] While presentations of these principles may differ in nuance and details, they are found in some form in virtually every contemporary presentation of tonal harmony. Of course, 'conventional wisdom' is not always correct, but recent research has shown that, indeed, these rules are largely followed in common-practice composition (Huron 2006; Temperley 2009).

What, then, of rock harmony? Is rock music also governed by general harmonic principles, and if so, what are they? The writings of three recent authors provide a representative picture of contemporary thinking on this issue.

Stephenson (2002) argues that rock harmony constitutes 'a new harmonic practice', quite distinct from that of common-practice tonality. His basic premise is that the normative harmonic principles of rock constitute a kind of 'mirror-image' to those of common-practice music. While ascending fourths, descending thirds and ascending seconds are the most common root motions in common-practice harmony, the normative motions of rock – according to Stephenson – are precisely the opposite: descending fourths, ascending thirds and descending seconds. (Stephenson refers to this latter practice as 'rock-standard' harmony.) Stephenson acknowledges, however, that many rock songs include examples of common-practice root motion or combinations of the two practices. Stephenson suggests also that many rock songs employ one of three harmonic systems or 'palettes': the natural-minor system (allowing major and minor triads native to the natural-minor scale), the major system (allowing major and minor triads built on Mixolydian scale degrees) and the chromatic-minor system (allowing major triads built on natural-minor scale degrees).

In the writings of Moore (1992, 1995, 2001), modality plays a central role. Moore argues that many rock progressions are confined to a single diatonic mode (with Ionian, Mixolydian, Dorian and Aeolian being most common); even those that are not can often be regarded as altered modal progressions. With regard to root motion, Moore suggests that motion by fourths (both ascending and descending) is primary in rock, while motion by seconds or thirds is rarer; he notes that many root progressions involve 'cyclic' repetitions of a single interval, such as ascending or descending fourth progressions. Moore also argues, however, that harmony in rock is frequently 'non-functional', in the sense that chords often do not evoke strong expectations for specific continuations. With regard to larger-scale patterns, Moore invokes the common-practice notion of a 'period' – two phrases in a 'question/answer' relationship, with the first ending on non-tonic harmony (usually on V) and the second on tonic; Moore argues that period structures are commonplace in rock, but notes that non-periodic ('open-ended') harmonic structures are frequent as well, often consisting of a short repeated harmonic pattern.

Everett (2004, 2008) rejects the view of rock as a unified harmonic language, stating that 'there is no single monolithic style of rock harmony' (2004, paragraph 37). Everett posits six distinct tonal systems for rock, characterised by varying degrees of adherence to common-practice norms; the systems range from full-blown common-practice tonality, through modal and blues-based styles, to highly chromatic music in which common-practice principles have little relevance. Everett's approach is strongly influenced by Schenkerian theory; under this view, tonality is seen as a hierarchical system wherein events elaborate other events in a recursive fashion. This has important implications for harmony; in particular, Everett often explains surface harmonies in contrapuntal terms, that is as neighbouring or passing events or as part of long-range linear patterns. Everett's analyses generally follow the normal Schenkerian practice of positing a I-V-I progression at the deepest level. But Everett emphasises that this should not necessarily be taken to imply that such a structure is literally present; in many cases, he argues, the Schenkerian approach is useful precisely as a way of showing how rock songs deviate from common-practice norms.

We do not wish to oversimplify these authors' views; no doubt they are more complex and nuanced than our brief summaries might suggest. Nor do we wish to dwell on the amount of agreement or disagreement between them. We do, however, wish to make two general points. First, it is notable that all three authors view rock harmony, to a large extent, in terms of its adherence to or rejection of common-practice norms. Even Stephenson's 'rock-standard' is defined in relation to common-practice harmony (as the opposite of it). This approach is totally understandable. Common-practice harmonic theory provides an elegant and empirically robust set of principles with which most writers and readers of rock scholarship are familiar; it therefore provides a natural point of reference to which other styles may be compared. There is a danger, however, that this perspective may cause us to overlook other kinds of logic operative in rock that are best understood neither as expressions or rejections of common-practice norms, but simply on their own terms. We hasten to add that our own backgrounds are heavily influenced by common-practice theory as well, and we may be as biased by its influence as anyone else. But we suggest that, in view of this potential bias, a more 'data-driven' approach to rock harmony may be desirable, an approach in which the music is allowed to speak for itself.

A second point about these authors' views is that the harmonic principles they propose, to the extent that they can be considered principles at all, are of an extremely loose and unrestrictive nature. Both Everett's and Stephenson's theoretical frameworks, in rather different ways, essentially allow for a range from complete accordance with common-practice norms to complete rejection of them. But this would seem to imply that almost anything can happen in rock harmony. Is this really the case? If so, it is in stark contrast to the situation in common-practice harmony, where certain progressions, such as II-I or V-IV, are generally considered incorrect, or usable only under highly constrained circumstances. Perhaps rock harmony, too, is guided by strong and restrictive principles that have not yet been observed. One might say that we are guilty here of exactly the kind of imposition of common-practice thinking that we warned against in the previous paragraph, and perhaps we are. But again, we would argue that such questions can be best resolved by adopting a more objective approach, in which the facts of harmonic practice in rock can be observed in a direct and unbiased way.

One powerful way of achieving the 'objective approach' that we have recommended is through statistical corpus analysis. Under this approach, statistical information is gathered from the music itself. While statistical corpus methods have been a part of music research for several decades at least (see, for example, Budge 1943), the last decade has seen greatly increased activity in this area – no doubt due to the general rise of scientific, empirical approaches in music research and also to the new possibilities afforded by computer technology. Especially notable is the work of David Huron (2006). Huron and his colleagues have applied corpus analysis to investigate a range of issues in music theory and music psychology. To some extent, this work has provided empirical support for longstanding principles of music theory (for example, the preference for imperfect over perfect consonances in two-part counterpoint); in other cases, it has overturned conventional ideas (such as the notion of 'post-skip reversal' as a principle of melodic structure) or revealed entirely new regularities of musical structure (such as the asymmetry between diminuendos and crescendos). The work of Huron and others has shown that corpus analysis can make useful contributions to a wide variety of musical questions, providing objective answers in place of conjecture and guesswork, and frequently yielding new and unexpected insights.

In what follows, we present a statistical corpus analysis of rock harmony. We believe this project is of interest in two ways. First of all, it is one of the first large-scale, systematic corpus analyses of popular music, and thus raises some fundamental general issues about how corpus analysis should be done in this area. Secondly and more specifically, it provides a body of empirical evidence regarding the norms and regularities of harmonic practice in rock, which we believe may provide a useful foundation for future explorations of the topic.

Methodological issues

Our assumption is that rock songs have an underlying harmonic structure, which is known (at least on an unconscious level) to the creators of the songs and also to experienced listeners to the style. We further assume that the creation of these structures is guided by certain underlying principles; our aim is to identify these principles. Thus our primary interest is in a kind of knowledge (again, perhaps largely unconscious) that underlies the creation and perception of rock. While we do not have direct access to this knowledge, we can use actual musical objects (rock songs) as indirect evidence of it; observing patterns in the harmonic structures of rock songs may tell us something about the principles involved in their creation. Statistical corpus analysis is a way of achieving this goal. (For further discussion of this approach, see Temperley 2007b.)

Two significant methodological issues arise here. One is how the corpus is to be selected; the second is how harmonic information is to be extracted from it.

Our aim in this study is to examine the harmonic structure of rock. However, 'rock' proves to be a problematic term. Roughly speaking, it seems to be used in two different senses. In one sense, rock is a broad and eclectic category incorporating a wide spectrum of late 20th century Anglo-American popular music. For example, VH1's list of the '100 Greatest Songs of Rock & Roll' (1999) includes not only groups like the Rolling Stones, Aerosmith and the Police, which would surely count as rock under any definition, but also soul/R&B standards like 'Respect', the disco hit 'Stayin' Alive' and the MOR (middle-of-the-road) pop classic 'We've Only Just Begun'. Most books on rock reflect a similarly broad construal of the term.[2] But there is a more limited sense of the term as well. Data on record sales from the Recording Industry Association of America (RIAA) reflect a much narrower definition of rock than that reflected in the VH1 list, distinguishing it from other categories such as pop, R&B/urban and rap/hip-hop.[3] Similarly, the online music database Allmusic.com (2010)

https://doi.org/10.1017/S026114301000067X Published online by Cambridge University Press

lists 'pop/rock' as a separate category from 'R&B'; 'Respect' is in the latter category, and the term 'pop/rock' itself suggests that some 'pop' music is not 'rock'.

Clearly, there is some disagreement as to what is rock and what is not. One might surmise that rock, or rather the 'rock song', is not a discrete, well defined category, but a loosely defined schema, of which some songs are very clear-cut instances and others are more borderline and ambiguous. It is also possible, of course, that different people simply have different understandings of the term. In light of this, and given the aim of the study to employ empirical, statistical methods, it seems most sensible to adopt a statistical approach to the definition of rock as well, that is, basing it on the opinions of many individuals.

One way to do this is by using what might generically be called 'greatest' lists. Music media outlets (magazines, radio stations and cable channels) often compile lists of the 'greatest' songs in rock and other styles. In some cases, these lists are generated solely by the magazine editors or disc jockeys themselves. As a result, the inclusion or exclusion of a song depends entirely on the opinions of only a few individuals. In other cases, however, these lists are created by compiling a large number of opinions, for example, votes by radio listeners or polls of people in the music business. If someone includes a song in their list of the 'greatest rock songs', this presumably indicates that they consider it to be rock. Whether or not the songs on the list are in fact the *greatest* rock songs (whatever that might mean) is not important for present purposes; all that matters is that they are rock songs. This methodology is not foolproof; a song that was rated, for example, as number one by half the participants but was not considered to be rock by the other half (and thus not included in their lists) might well be included in the final list, despite its ambiguous status as rock. But as far as we know, this is the best available method, using currently available data, of creating a corpus of songs that are generally considered to be rock.

The list we chose was *Rolling Stone* magazine's '500 Greatest Songs of All Time' (2004). This is one of the few lists we have found that purports to reflect 'rock' in a general sense, without any stylistic modifiers (e.g. 'hard rock') or limitations by era or decade.[4] While the title of the *Rolling Stone* list does not indicate a specific focus on rock, the introductory text of the article makes its focus clear; this text states that the list is 'a celebration of the greatest rock & roll songs of all time' and that the voters were asked to 'select songs from the rock & roll era'.[5] Another significant factor in our selection of this list was the large number of participants polled: 172 'rock stars and leading authorities' were asked to provide their 50 top songs, and these song selections were combined using a 'numerical weighting system'. Although the list was compiled in 2004, it is somewhat weighted towards the earlier decades of rock; 277 of the 500 songs are from before 1970 and 203 are from the 1960s alone. To balance the corpus more evenly across rock's history, we created a smaller list consisting of the top 20 songs from each decade: the 1950s, 1960s, 1970s, 1980s and 1990s. (The *Rolling Stone* list contains only four songs from 2000 or later, and we did not include these.) This collection of 100 songs, which we call the 'RS 5 × 20 corpus', is shown in Table 1.

The second challenge we confronted was how to extract the needed harmonic information from each song. We take a rock song to be defined by a particular recording. Recordings do not, of course, indicate harmonic structure in any obvious or explicit way; this structure has to be identified somehow. This problem is hardly unique to rock; in common-practice music, a piece is defined by a notated score rather than a recording, but scores do not usually carry explicit harmonic information

Table 1. The RS 5 × 20 corpus (the top 20 songs in each decade from Rolling Stone*'s '500 Greatest Songs of All Time').*

Rank	Decade	Artist	Song
7	1950s	Chuck Berry	Johnny B. Goode
10		Ray Charles	What'd I Say
18		Chuck Berry	Maybellene
19		Elvis Presley	Hound Dog
30		Johnny Cash	I Walk the Line
39		Buddy Holly	That'll Be the Day
43		Little Richard	Tutti Frutti
45		Elvis Presley	Heartbreak Hotel
56		Little Richard	Long Tall Sally
61		Jerry Lee Lewis	Whole Lotta Shakin' Goin On
62		Bo Diddley	Bo Diddley
67		Elvis Presley	Jailhouse Rock
73		Eddie Cochran	Summertime Blues
77		Elvis Presley	Mystery Train
81		Fats Domino	Blueberry Hill
90		The Five Satins	In the Still of the Nite
94		Little Richard	Good Golly, Miss Molly
95		Carl Perkins	Blue Suede Shoes
96		Jerry Lee Lewis	Great Balls Of Fire
97		Chuck Berry	Roll Over Beethoven
1	1960s	Bob Dylan	Like a Rolling Stone
2		The Rolling Stones	Satisfaction
5		Aretha Franklin	Respect
6		The Beach Boys	Good Vibrations
8		The Beatles	Hey Jude
11		The Who	My Generation
12		Sam Cooke	A Change Is Gonna Come
13		The Beatles	Yesterday
14		Bob Dylan	Blowin' in the Wind
16		The Beatles	I Want to Hold Your Hand
17		The Jimi Hendrix Experience	Purple Haze
22		The Ronettes	Be My Baby
23		The Beatles	In My Life
24		The Impressions	People Get Ready
25		The Beach Boys	God Only Knows
26		The Beatles	A Day in the Life
28		Otis Redding	(Sittin' on) the Dock of the Bay
29		The Beatles	Help!
32		The Rolling Stones	Sympathy for the Devil
33		Ike & Tina Turner	River Deep - Mountain High
3	1970s	John Lennon	Imagine
4		Marvin Gaye	What's Going On
20		The Beatles	Let It Be
21		Bruce Springsteen	Born to Run
27		Derek and the Dominos	Layla
31		Led Zeppelin	Stairway To Heaven
37		Bob Marley & The Wailers	No Woman, No Cry
46		David Bowie	Heroes

Continued

https://doi.org/10.1017/S026114301000067X Published online by Cambridge University Press

Table 1. Continued

Rank	Decade	Artist	Song
47		Simon and Garfunkel	Bridge over Troubled Water
49		The Eagles	Hotel California
53		Sex Pistols	Anarchy in the U.K.
60		Al Green	Let's Stay Together
68		Bob Dylan	Tangled Up in Blue
74		Stevie Wonder	Superstition
86		Bruce Springsteen	Thunder Road
92		The Ramones	Blitzkrieg Bop
98		Al Green	Love and Happiness
103		Donna Summer	Hot Stuff
104		Stevie Wonder	Living for the City
117		Al Green	Take Me to the River
15	1980s	The Clash	London Calling
51		Grandmaster Flash & the Furious Five	The Message
52		Prince and the Revolution	When Doves Cry
58		Michael Jackson	Billie Jean
66		Bob Marley & The Wailers	Redemption Song
84		The Police	Every Breath You Take
93		U2	I Still Haven't Found What I'm Looking For
108		Prince	Little Red Corvette
131		U2	With or Without You
143		Prince	Purple Rain
160		Public Enemy	Bring the Noise
165		Tracy Chapman	Fast Car
177		Tom Petty	Free Fallin'
179		Joy Division	Love Will Tear Us Apart
187		AC/DC	Back in Black
196		Guns N' Roses	Sweet Child O' Mine
201		New Order	Bizarre Love Triangle
212		Prince	1999
214		Neil Young	Rockin' in the Free World
228		The Clash	Should I Stay or Should I Go
9	1990s	Nirvana	Smells Like Teen Spirit
36		U2	One
162		Sinead O'Connor	Nothing Compares 2 U
169		R.E.M.	Losing My Religion
200		Beck	Loser
256		Radiohead	Paranoid Android
259		Jeff Buckley	Hallelujah
286		Pavement	Summer Babe
331		Bonnie Raitt	I Can't Make You Love Me
346		Dr. Dre and 2Pac	California Love
353		Eric Clapton	Tears in Heaven
376		Radiohead	Fake Plastic Trees
382		The Verve	Bitter Sweet Symphony
399		Metallica	Enter Sandman
406		R. Kelly	I Believe I Can Fly
407		Nirvana	In Bloom

Continued

https://doi.org/10.1017/S026114301000067X Published online by Cambridge University Press

Table 1. Continued

Rank	Decade	Artist	Song
419		Dr. Dre	Nuthin' But a 'G' Thang
445		Nirvana	Come As You Are
455		Nirvana	All Apologies
475		The Beastie Boys	Sabotage

either. It is well known in common-practice theory that harmonic analysis is somewhat subjective; two analysts (even using the same labelling system) will not necessarily analyse a piece in the same way. We assumed that this would be the case in rock as well. While published notated arrangements (e.g. piano/vocal scores) are available and provide harmonic symbols for most songs, these arrangements are generally not created by the writers of the song but rather by professional arrangers; these publications therefore constitute only one opinion about a song's harmonic structure which (in our view) need not be given priority over any other. In addition, notated arrangements usually do not provide complete harmonic analyses. Specifically, only 'absolute' chord labels (for example, 'Cm' or 'G7') are provided, not the relationship of these chords to a tonal centre; identifying relative chord labels can be a non-trivial issue, as we will see.

In light of all this, we decided the best solution was to do the harmonic analyses ourselves. Again, harmonic analysis is subjective, and it is possible that our analyses were affected by idiosyncracies or biases (such as an unconscious desire to prove a particular theory). To mitigate this problem, both authors independently analysed every song in the corpus. This served two purposes. First, by comparing our analyses, we were able to measure quantitatively the amount of agreement between them, thus giving some indication as to the subjectivity of harmonic analysis in rock. Secondly, by combining our analyses, we produced a dataset that was not too greatly influenced by the idiosyncracies of any one individual. (While both of us have considerable training in common-practice theory, our backgrounds are otherwise quite different: de Clercq [TdC] has extensive experience as a rock musician, while Temperley [DT] was formerly an art-music composer and an accompanist for dance and musical theatre.) Of course, both of these aims (measuring agreement and eliminating the effect of idiosyncracies) would be still better served if the songs were analysed by more than two people, but even two sets of analyses seemed to us to be significantly better than one.

Analysing the harmony

In analysing the songs in the RS 5 × 20 corpus, our aim was to extract the harmonic structures, with the ultimate goal of finding patterns or regularities in those structures. We assumed that, for the most part, these patterns would best be characterised in relative terms, that is, relative to a tonal centre. Again, this is the usual assumption in common-practice music, where conventional harmonic patterns are generally described not in absolute terms (e.g., G major to C major) but rather in terms of

Roman numerals, that is, roots in relation to the key (e.g. V to I). Roman numeral analysis seemed to be the most appropriate notational system for our analyses as well.

We devised a simple syntax for our analyses. Here is an example:

[C] I ii7 | V I | . IV | I ||

The 'C' in brackets at the beginning indicates the tonal centre. It is often difficult to categorise rock songs in the common-practice tonal systems of 'major' or 'minor'; many songs adhere to other modes (such as Mixolydian), freely mix triads from the major and minor modes, or even combine elements of major and minor simultaneously (e.g. a minor third in the melody over a major triad). However, the tonal centre of a song can usually be identified quite unambiguously (though not always, as we will discuss). (Here we will use 'key' simply to mean 'tonal centre' in a broad sense, with no implication of common-practice tonality.)

Following the key symbol, Roman numerals are organised into measures delineated by vertical lines (which indicate barlines). A measure may contain multiple harmonic symbols, in which case our notational system assumes that the measure is evenly divided among them. The dot '.' can be used to indicate a segment with no change of harmony; for example, the dot in the third measure in the example above indicates that the previous I harmony extends through the first half of the measure. More complicated harmonic rhythms may be notated with a combination of Roman numerals and dots, such as | I V . . |, in which I occupies the first quarter of the measure while V occupies the remainder. Two barlines with nothing intervening indicate that the previous harmony continues through that entire measure.

Our use of Roman numerals follows convention (or at least, one widely used convention). Capital letters indicate major triads; lower-case letters indicate minor triads. Roots are assumed to be major-mode scale degrees unless otherwise indicated; for example, 'bVI' indicates a major triad on the flattened sixth degree. (This should not be taken to indicate any theoretical commitment to major mode as the 'primary' mode of rock; it was simply a notational convenience.) Following the Roman numeral, we used a variety of symbols to indicate inversions (e.g. 'I6', 'I64', 'I42'), extensions ('i7' for a minor seventh chord, 'I7' for a major seventh), and other qualities (e.g. 'o' for a diminished triad); we also allowed applied chords ('V/V'). However, we did not attempt to fully standardise our use of symbols following the Roman numerals; for the most part, the statistical analyses presented below consider only the Roman numerals themselves. (Applied chords are converted into Roman numerals relative to the key, e.g. V/V becomes II.)

Most rock songs contain a good deal of repetition in their harmonic structures. Rather than explicitly notating the entire harmonic progression of each song, it seemed more efficient to represent harmonic structures in a hierarchical fashion. We devised a system in which a higher-level element can be defined to represent a series of lower-level elements. The system is similar to phrase structure grammars of the kind commonly seen in linguistics, in which 'non-terminal' symbols (such as 'noun phrase' or 'verb phrase') expand to other non-terminals and eventually to 'terminal' symbols (words); in this case, the terminals are harmonies, and the non-terminals are larger units (e.g. phrases or sections). As an example, Figure 1a shows TdC's analysis of 'Hey Jude' by the Beatles. We indicate the highest level of each song as 'S' (for song), which can be seen in the last line of this analysis. The definition of a non-terminal may include terminals, non-terminals, or a combination.

```
Vr: I | V | V7 | I | IV | I | V | I |
A: V7 | I I42 | vi vi42 | V6/V V/V | V |
Br: [Bb] $A*2 [2/4] V7 | [4/4] [F] V | V7 |
Ou: I | bVII | IV | I |
S: [F] $Vr*2 $Br $Vr $Br $Vr I | $Ou*18
```

Figure 1a. TdC's reduced analysis of 'Hey Jude'.

```
Vr: I | V | | I | IV | I | V | I |
BrP: V7/IV | IV I6 | ii vi6 | V6 V | I |
Br: $BrP $BrP [2/4] I | [4/4] V | |
CP: I | bVII | IV | I |
Fadeout: $CP*18
S: [F] $Vr $Vr $Br $Vr $Br $Vr I | $Fadeout
```

Figure 1b. DT's reduced analysis of 'Hey Jude'.

(When a non-terminal is used in the definition of another non-terminal, it is preceded by the '$' sign to distinguish it from a chord symbol.) In Figure 1a, for example, the definition of 'S' contains non-terminals such as $Vr (verse) and $Br (bridge) as well as the terminal symbols 'I |' (a measure of tonic) near the end. (Our syntax allows multiple iterations of a non-terminal to be listed successively, e.g. $Vr $Vr, or more succinctly, as seen here, as $Vr*2, which indicates two iterations of the verse.) The non-terminals used in the definition of 'S' are then given their own definitions, which again may consist of any combination of terminals and non-terminals. The system is recursive; there can be any number of levels of elements defined as other elements. In practice, we rarely found a need for more than four hierarchical levels.

The example of TdC's 'Hey Jude' analysis also shows how our notation represents metre and tonal centre. Time signatures are indicated by symbols such as [4/4] or [3/4]. Since the vast majority of songs in the corpus are in 4/4, we treated this metre as a default for all songs. Thus the only time signature indications necessary in 'Hey Jude' were at the end of the bridge section ('Br'), where a measure of 2/4 is inserted. With regard to key, TdC hears a modulation to (or perhaps prolonged tonicisation of) the subdominant during the majority of the bridge section; thus, while the main tonal centre is F (indicated in the 'S' definition), the 'Br' definition indicates a shift to Bb. A key or metre symbol applies to any nonterminals that follow; for example, the [Bb] symbol in Figure 1a indicates that the 'A' unit should be interpreted in Bb. Once a non-terminal is exited, however, any local key or metre statements are reset back to the higher-level statements.

We then wrote a computer programme that takes a 'reduced' analysis such as Figure 1a and expands it, replacing 'S' with its definition and then rewriting other symbols recursively to produce an 'expanded' analysis. The expanded analysis consists of chords only (no non-terminals), and every chord in the song is explicitly shown. Figure 2 shows the expanded analysis generated from TdC's analysis of 'Hey Jude'.

A number of issues arose in doing these analyses. There were many points of ambiguity. We found we were often uncertain whether something was a major or minor triad. In many cases only the root and fifth of a chord are heard (i.e. 'power chords' which are open fifths on an electric guitar); in other cases (as noted earlier), major and minor thirds above the root may both be heard. It was also often difficult to decide whether to include sevenths and other extensions. There were ambiguities

[F] I | V | V7 | I | IV | I | V | I | | V | V7 | I | IV | I | V | I | [Bb] V7 | I I42 | vi vi42 | V6/V V/V | V |
V7 | I I42 | vi vi42 | V6/V V/V | V | [2/4] V7 | [4/4] [F] V | V7 | I | V | V7 | I | IV | I | V | I | [Bb]
V7 | I I42 | vi vi42 | V6/V V/V | V | V7 | I I42 | vi vi42 | V6/V V/V | V | [2/4] V7 | [4/4] [F] V |
V7 | I | V | V7 | I | IV | I | V | I | | | bVII | IV | I | | bVII | IV | I | | bVII | IV | I | | bVII | IV | I | | bVII
| IV | I | | bVII | IV | I | | bVII | IV | I | | bVII | IV | I | | bVII | IV | I | | bVII | IV | I | | bVII | IV | I | |
bVII | IV | I | | bVII | IV | I | | bVII | IV | I | | bVII | IV | I | | bVII | IV | I | | bVII | IV | I | | bVII | IV
| I |

Figure 2. TdC's analysis of 'Hey Jude' (expanded).

regarding key and root labels as well; we will discuss these further below. Some ambiguities of metre also arose; in some cases, it seemed that a particular rhythmic unit could be considered as the measure, but that a unit half as long or twice as long would also be plausible. Another problem was that many songs do not have a definite ending but rather fade out on a repeating harmonic pattern; in such cases, we agreed to include all iterations of the pattern up to and including the last iteration that could be heard in its entirety.

We each analysed all 100 songs in the RS 5 × 20 corpus in this fashion.[6] We analysed the songs entirely by ear without consulting other sources (e.g. lead sheets, online transcriptions, etc.). We also did not consult with each other. When the process was finished, we compared our analyses. We were forced to resolve certain differences in order for our harmonic analyses to be properly compared. For example, when we chose main keys that were a half-step apart (because the actual key was somewhere 'between the cracks'), we altered one or another of the analyses so that they agreed. We also resolved differences in metre, such as cases where we had chosen measures of different lengths. When we encountered differences of any kind that were due to unintentional errors in one analysis or another, these were corrected. We did not, however, resolve intentional, musically substantive differences in harmonic analysis, since part of our aim was to evaluate how much agreement there was in this regard.

Figure 1b shows DT's analysis of 'Hey Jude'. It is instructive to compare this to TdC's analysis in Figure 1a. Some differences are completely superficial; for example, both analyses define the opening five measures of the bridge section as a unit, but they use different labels for this unit ($A versus $BrP). Other differences are more significant; TdC hears the bridge as centred on Bb while DT does not posit any modulation away from the original key of F (see Example 1). As a result, the analyses generate

Example 1. Selected instruments from the beginning of the middle section in 'Hey Jude'. (Words and Music by John Lennon and Paul McCartney, Copyright © 1968 Sony/ATV Music Publishing LLC, Copyright Renewed, All Rights Administered by Sony/ATV Music Publishing LLC, 8 Music Square West, Nashville, TN 37203, International Copyright Secured, All Rights Reserved, Reprinted by permission of Hal Leonard Corporation)

https://doi.org/10.1017/S026114301000067X Published online by Cambridge University Press

different sets of relative roots for this passage. More agreement can be seen here with regard to absolute roots, but some disagreement still remains. While TdC finds a single harmony in the second measure of the bridge with a changing inversion I-I42 (implying a Bb harmony throughout), DT identifies a true change of harmony IV-I6 (implying Bb to F); a similar disagreement is found in the following measure.

This latter difference between our analyses of 'Hey Jude' is representative of the kind of ambiguity that often arose in our analyses. Both of our analyses seem plausible. The pitch-classes of the background vocals in the second half of the bridge's second measure are clearly A-C-F (supported by an A in the bass), as shown in Example 1; thus the harmony at this point seems to move to a first-inversion F-major harmony (as in DT's analysis). (The D of the melody could be explained as a purely melodic note, independent of the changing harmonies beneath – an example of what some have called 'melodic-harmonic divorce' [Moore 1995, p. 189; Temperley 2007a].) Yet at the same time, the melody for this entire measure clearly outlines a Bb-major triad, and it could be argued that this expresses the underlying harmony (as in TdC's analysis); by this view, the A and C of the accompaniment are simply passing tones. The issue, then, is whether to regard the A and C as expressing a true harmony or as purely linear elaborations; many of the disagreements between our analyses were of this nature.

As a final step, each expanded analysis was converted to what we call a 'chord list'. The beginning of the chord list for TdC's analysis of 'Hey Jude' is shown in Figure 3. The analysis is mapped on to a timeline of measures; measure 1 starts at time 0.0, measure 2 starts at time 1.0, and so on. Each harmony is shown on its own line, along with the timepoints at which it begins and ends. Chord changes within a measure can be indicated with fractional timepoints (e.g. 22.50). Following the timepoints and the Roman numeral label are four integers. The first integer indicates what we call the 'chromatic relative root', that is, the root as a chromatic interval above the tonic: I (or i) is 0, bII (or #I) is 1, II is 2, and so on. The second integer indicates the 'diatonic relative root' – essentially, the actual Roman numeral number. (I is 1, bII and II are both 2, and so on.) The third integer indicates the key in pitch-class terms (assuming the conventional mapping: C = 0, C#/Db = 1, etc.), and the fourth indicates 'absolute root' – the pitch-class of the root in absolute terms (using the same pitch-class

0.00	1.00	I	0	1	5	5
1.00	3.00	V	7	5	5	0
3.00	4.00	I	0	1	5	5
4.00	5.00	IV	5	4	5	10
5.00	6.00	I	0	1	5	5
6.00	7.00	V	7	5	5	0
7.00	9.00	I	0	1	5	5
9.00	11.00	V	7	5	5	0
11.00	12.00	I	0	1	5	5
12.00	13.00	IV	5	4	5	10
13.00	14.00	I	0	1	5	5
14.00	15.00	V	7	5	5	0
15.00	16.00	I	0	1	5	5
16.00	17.00	V7	7	5	10	5
17.00	18.00	I	0	1	10	10
18.00	19.00	vi	9	6	10	7
19.00	20.00	V6/V	2	2	10	0
20.00	22.00	V	7	5	10	5
22.00	23.00	I	0	1	10	10

Figure 3. Beginning of chord list for TdC's analysis of 'Hey Jude' (see Figure 2).

https://doi.org/10.1017/S026114301000067X Published online by Cambridge University Press

mapping, C = 0). (The absolute root is not indicated explicitly in our analyses, but can be inferred from the Roman numeral and the key: for example, the absolute root of IV of C must be F.) The four integer columns ignore everything about harmonies except for root and key; if two adjacent harmonies differ only in inversion or extension, they are collapsed into a single 'chord span'. (This occurs, for example, with the V6/V-V/V motion in m. 20 (timepoint 19.00) of 'Hey Jude'; compare Figures 1a and 3.) It was these chord lists that provided the data for the statistical analyses presented in the next section.

Patterns in the data

Having created a corpus of harmonic analyses of rock songs, we could address our actual goal, which was to examine what kinds of regularities, patterns and general principles of harmonic structure emerge from the data. While our investigation of this is ongoing, in what follows we present some of our preliminary results.

Before addressing this issue, we must first consider the issue of agreement. To what extent is the harmonic analysis of rock subjective and susceptible to differences of opinion between analysts? We suggested earlier that we could examine this quantitatively by comparing TdC's and DT's analyses of the RS 5 × 20 corpus. Given chord lists of the kind described in the previous section, this comparison could be done quite easily. From the chord lists, one can identify the total time for all 100 songs (using the measure units indicated by the timeline) as well as the amount of time that the two analyses of the songs are in agreement; the ratio of 'time in agreement' to 'total time' gives a measure of the level of agreement between the analyses. One issue was how the chord lists should be compared. We decided not to compare actual harmonic symbols; many of the differences between these concerned things such as major *vs.* minor and triad *vs.* seventh, which are undoubtedly often ambiguous and subjective (as discussed earlier), but seemed less important to us than actual root judgements. We decided to compare two things: chromatic relative root (the first integer column) and absolute root (the fourth integer column). If the two analyses agree on the chromatic relative root of a segment, they presumably agree on both the key and also on the absolute root (since the absolute root follows from the key and the relative root).[7] It is possible, however, that two analysts might agree on absolute root while disagreeing on chromatic relative root – for example, if they disagreed on the underlying key. Therefore, we examined agreement in both chromatic relative root and absolute root.

With regard to chromatic relative root, the level of agreement between our analyses was 92.4 per cent; with regard to absolute root, it was 94.4 per cent.[8] We will focus here on relative root agreement. On 39 of the 100 songs, there was 100 per cent agreement between the analyses with regard to relative root. (One song, Public Enemy's 'Bring the Noise', was judged by both of us not to contain any triadic harmony, and was therefore not included in any further analyses.) Of the remaining songs, 39 displayed a level of agreement of 90 per cent or greater. On nine songs, the level of agreement was below 70 per cent. Inspection of these latter cases showed that the disagreements were due to several factors. In some cases, the two analyses notated part or most of the song in different keys; the bridge of 'Hey Jude', discussed above, is one example of this. (Overall, our analyses were in agreement with regard to key 97.3 per cent of the time.) In other cases, a basic recurring progression in a song was analysed in different ways. In the Rolling Stones song 'Satisfaction', for

Table 2. Distribution of chromatic roots in the RS 5 × 20 corpus.

Root	Instances	Proportion of total	Song instances
I	3,058	0.328	99
bII	46	0.005	5
II	336	0.036	39
bIII	240	0.026	18
III	174	0.019	23
IV	2,104	0.226	90
#IV	23	0.003	4
V	1,516	0.163	88
bVI	372	0.040	21
VI	674	0.072	39
bVII	748	0.081	37
VII	38	0.004	7

example, the same two-measure progression (underlying the famous guitar riff) occurs throughout the introduction, chorus and fadeout; DT analysed this progression as 'I | IV |' while TdC analysed it as 'I . . IV | bVII64 . . IV |'.[9] Overall, however, the level of agreement between our analyses suggests that there is a good deal of common ground in how harmonic analysis is done and that our analyses can therefore be taken as a fairly reliable body of data for further investigation.

Once we had both analysed the entire corpus, it was necessary to combine our analyses in some way to allow aggregate analysis of the data. Where our analyses disagreed, it seemed to us that both analyses represented valid points of view, and we thought these differences should be preserved in the data. Thus, for each of the statistics reported below, we calculated the statistic separately for TdC's analyses and DT's analyses, and then took the average of the two; it is these averages that are reported here.[10]

The first thing we examined was the overall distribution of chromatic relative roots. This is shown in Table 2. The first column shows the sheer number of occurrences of each root; the second column shows the proportion of the total. (As noted earlier, multiple adjacent chords of the same root are collapsed into a single chord.) Not surprisingly, I is the most frequent; next most frequent is IV, followed by V, bVII and VI. (We use only capital Roman numerals here, though it should be borne in mind that these categories include both major and minor triads and all other chords of the same root.) These five roots account for 87 per cent of all harmonies in the corpus. Even this preliminary data clearly shows departures from common-practice norms – notably the high incidence of bVII, which is quite rare in common-practice music (both in major and minor mode) (Temperley 2009). The rightmost column of Table 2 shows the number of songs in which each chord occurs. The ranking here is somewhat different from that obtained from the raw counts in the leftmost column. The status of II is of particular interest in this regard; it appears that II is used in roughly as many songs as bVII and VI, but tends to be used less repetitively, thus resulting in a lower overall count.

We next considered chord transitions, that is, moves from one root to another. The data is shown in Table 3. (The data reflects only transitions within a single key; no transition is recorded for moves from one key to another.) It is interesting to

Table 3. Chord transitions in the RS 5 × 20 corpus. Cells indicate the number of occurrences from one chord (the 'antecedent') to another (the 'consequent').

Ant \ Cons	I	bII	II	bIII	III	IV	#IV	V	bVI	VI	bVII	VII
I	0	25	132	94	44	1052	2	710	104	302	470	16
bII	31	0	0	0	2	0	0	0	0	0	0	12
II	120	1	0	2	20	58	0	97	0	24	10	0
bIII	50	6	6	0	0	64	2	2	67	0	41	0
III	16	0	39	0	0	46	0	6	0	60	3	4
IV	1,162	14	30	98	45	0	4	514	57	72	90	4
#IV	7	0	0	6	0	10	0	0	0	0	0	0
V	788	0	36	6	17	392	4	0	6	191	48	0
bVI	208	0	1	20	0	22	6	22	0	10	78	0
VI	144	0	87	0	32	260	0	124	21	0	3	0
bVII	386	0	0	11	2	188	2	26	114	6	0	0
VII	18	0	0	0	12	0	4	0	0	3	0	0

consider this data in light of the norms of common-practice harmony. In that case, as mentioned earlier, the general rule is that pre-dominants (IV and II) move to dominants (primarily V), dominants move to tonics, and tonics can move freely to any chord. There is little evidence of these norms in the data in Table 3. The leftmost column indicates the distribution of what we might call 'pre-tonic' chords; these are chords moving to (immediately preceding) the tonic. It can be seen that the most frequent chord to precede the tonic is IV; V is a strong second and bVII is third. The top row shows the distribution of chords that immediately follow I: 'post-tonic' chords. This appears quite similar to the pre-tonic distribution, featuring the same top three chords in the same order: IV, V and bVII. This again points to a very strong contrast with common-practice harmony, where harmonic motion around the tonic is highly asymmetrical: I often moves to ii or IV but is rarely approached from them.

It might be observed that the 'pre-tonic' distribution (chords approaching I), the 'post-tonic' distribution (chords approached from I), and the overall distribution of roots (excluding the tonic) all have certain features in common: in each case, the most common chord is IV, followed by V and bVII. In light of this data, one might conclude that rock is not governed by rules of 'progression' at all; rather, there is simply an overall hierarchy of preference for certain harmonies over others, regardless of context. (This brings to mind Moore's [2001] comments about the 'non-functional' nature of rock harmony.) Table 4 allows us to examine this view more closely. This table shows the overall distribution of relative roots, along with their pre-tonic and post-tonic distributions in proportional terms. (The tonic harmony has been excluded since it can be neither pre-tonic nor post-tonic; thus the numbers reflect the frequency of each chord as a proportion of the total excluding tonic.) While Table 4 does indeed show a strong similarity between the three distributions, some significant differences are brought out as well. In particular, while IV is the most common non-tonic harmony overall, its proportional frequencies as a post-tonic chord and (even more so) as a pre-tonic chord are even greater than its overall frequency. This suggests that IV is particularly favoured as a way of approaching or leaving

Table 4. Overall, pre-tonic and post-tonic chord distributions (excluding tonic).

Root	Overall	Pre-tonic	Post-tonic
bII	0.007	0.010	0.009
II	0.053	0.041	0.044
bIII	0.038	0.017	0.032
III	0.028	0.005	0.014
IV	0.336	0.396	0.356
#IV	0.004	0.002	0.001
V	0.241	0.269	0.240
bVI	0.059	0.071	0.036
VI	0.107	0.050	0.102
bVII	0.119	0.132	0.159
VII	0.006	0.005	0.005

the tonic. Positional preferences for other harmonies also emerge; V shows a slight bias for pre-tonic usage, but not for post-tonic usage; bVII tends somewhat towards post-tonic usage; and VI is generally *avoided* in pre-tonic position (its pre-tonic frequency is much less than its overall frequency).

Patterns of harmonic motion can also be described in intervallic terms. In common-practice music, conventional theory dictates that certain root patterns are preferred over others: ascending motion by fourths is especially normative (much more so than descending fourth motion); descending thirds are favoured over ascending thirds, and ascending seconds over descending seconds (Schoenberg 1969). Are these principles observed in rock as well? Table 5 shows the statistics from the RS 5 × 20 corpus. Chromatic root motions are shown in the left column; the right column groups them into diatonic intervals (e.g. grouping major and minor seconds together). It can be seen immediately that the norms of common-practice music do not hold. For each interval, the ascending and descending forms are roughly equal in frequency. The ascending perfect fourth is almost exactly as common as the descending perfect fourth; for other intervals, too, a similar pattern is seen. Regarding the frequency of different intervals, a striking pattern emerges when the intervals are represented on the 'circle of fifths' (or fourths); see Table 6. (Each interval can be represented as a certain number of fourth

Table 5. Root motions in the RS 5 × 20 corpus categorised by interval size.

Chromatic interval	Instances	Diatonic interval	Instances
+m2	113	+M/m2	1,497
+M2	1,384		
+m3	410	+M/m3	736
+M3	326		
+P4	2,266	+P4	2,266
+/−TT	20	+/−TT	20
−P4	2,220	−P4	2,220
−M3	412	−M/m3	866
−m3	454		
−M2	1,386	−M/m2	1,548
−m2	162		

Table 6. Root motions on the circle of fifths.

Interval	−m2	+M3	−m3	+M2	−P4	–	+P4	−M2	+m3	−M3	+m2	TT
Instances	162	326	454	1,384	2,220	–	2,266	1,386	410	412	113	20

motions. For example, two descending fourths make an ascending major second, e.g. C-G-D; and so on.) The frequency of intervals decreases in a very regular way as circle-of-fifths distance increases.

One might wonder if rock harmony reflects a preference for certain larger patterns, beyond simple transitions from one chord to another. To explore this, we examined the frequency of relative-root 'trigrams' (borrowing a term from computational linguistics): groups of three adjacent chords. Table 7 shows a subset of this data which we found to be of particular interest: three-chord patterns in which the final chord is tonic but the first two are not. (The second one cannot be tonic, since in that case the second and third would be regarded as a single chord.) The most common patterns feature IV or V as the pre-tonic chord, as we would expect from data presented earlier. However, in cases where V is the pre-tonic chord, the preceding (pre-pre-tonic) chord is overwhelmingly likely to be IV. By contrast, when IV is pre-tonic, three chords (V, bVII and VI) are all quite common as the preceding chord. This suggests that the pre-tonic V primarily occurs in a IV-V-I context, while IV has a broader variety of pre-tonic uses. Regarding lower-ranked trigrams, it should be noted that some of these are largely due to one or two songs; for example, bIII-bVI-I occurs primarily in Nirvana's 'Smells Like Teen Spirit'.

As noted earlier, we included 20 songs from each decade (the 1950s through the 1990s) in the corpus as a way of balancing it historically. This also allows us to examine the changes in harmonic practice across time. We present just one such analysis, showing the overall distribution of chromatic roots in each decade (Table 8). The most obvious pattern here is the contrast between the 1950s and later decades. The 1950s distribution is completely dominated by I, IV and V; these three chords account for more than 95 per cent of the total harmonic content. In the 1960s, by contrast, we see a somewhat broader distribution, quite similar to the overall distribution across all decades (shown in Table 2). This suggests that the harmonic language of rock (or

Table 7. Harmonic 'trigrams' ending in tonic (and not beginning in tonic) in descending order of frequency.

Trigram	Instances
IV V I	352
V IV I	292
bVII IV I	146
VI IV I	126
bVII bVI I	103
bIII bVI I	66
II V I	63
bVI bVII I	60
V VI I	42
IV bVII I	39

Table 8. Overall proportion of chromatic roots in each decade, 1950s–1990s.

	1950s	1960s	1970s	1980s	1990s
I	0.423	0.327	0.302	0.332	0.313
bII	0.001	0.000	0.004	0.008	0.008
II	0.004	0.074	0.038	0.005	0.050
bIII	0.000	0.009	0.032	0.028	0.040
III	0.007	0.040	0.027	0.002	0.017
IV	0.321	0.239	0.226	0.220	0.187
#IV	0.000	0.001	0.009	0.000	0.000
V	0.221	0.146	0.154	0.170	0.153
bVI	0.001	0.003	0.054	0.047	0.062
VI	0.011	0.072	0.063	0.081	0.100
bVII	0.007	0.084	0.089	0.108	0.064
VII	0.006	0.006	0.004	0.000	0.007

at least its harmonic vocabulary) largely matured in the 1960s and has not fundamentally changed since then. While there are some differences between later decades, it may be unwise to attribute too much weight to them considering the fairly small body of data (20 songs) available for each. It does seem significant, however, that the 'flat side' harmonies (bIII, bVI and bII) generally reflect an increase in frequency between the 1960s and later decades; this may reflect the rise in the 1970s of hard rock and heavy metal, styles in which these harmonies play a prominent role.

Aside from patterns of root motion, an additional issue in harmony is what we might generally call 'palette' (to use Stephenson's term). This refers to patterns of co-occurrence among chords, such that chord X is perhaps likely to co-occur in a song with chord Y, but unlikely to occur with chord Z. (By 'co-occur', we mean simply that the two chords occur in the same song, not that they are necessarily adjacent or even close together.) Larger groups of co-occurring chords may also be found. The obvious example of this is the major and minor modes of common-practice music, which have distinct scales and therefore distinct chords; the diatonic modal system is another example. As noted in our introduction, authors such as Moore, Stephenson and Everett all, in various different ways, endorse the grouping of rock songs by mode or palette. However, our ability to address this issue here is limited. For one thing, mode depends highly on chord quality (major versus minor I, for example), which is imperfectly represented in our data (as discussed further below). Mode also depends on purely melodic inflections; for example, a passage over a major I chord might be classified as Ionian or Mixolydian, depending on whether the raised or flattened 7th degree is present in the melody. However, modes are to some extent reflected in root patterns; for example, bVI occurs in Aeolian but not in Ionian, Mixolydian or Dorian. Thus our corpus can shed some light on the issue of mode, and we present one very preliminary exploration of it here.

Given the 99 songs in our corpus, we can define a 99-dimensional vector (i.e. a list of 99 numbers) for each of the 12 relative roots, '1' if the chord is present in the song at all and '0' if it is not. (We thus ignore the number of times the chord occurs in the song.) The correlation between two such vectors indicates the degree of similarity between them, and thus the degree to which chord X tends to occur in the same songs as chord Y. Correlating every root with every other root produces a 12×12

Table 9. Correlations between chord vectors.

	I	**bII**	**II**	**bIII**	**III**	**IV**	**#IV**	**V**	**bVI**	**VI**	**bVII**	**VII**
I	–	–	–	–	–	–	–	–	–	–	–	–
bII	0.000	–	–	–	–	–	–	–	–	–	–	–
II	0.000	−0.062	–	–	–	–	–	–	–	–	–	–
bIII	0.000	0.230	−0.113	–	–	–	–	–	–	–	–	–
III	0.000	−0.035	**0.571**	−0.140	–	–	–	–	–	–	–	–
IV	0.000	−0.078	0.079	0.057	0.054	–	–	–	–	–	–	–
#IV	0.000	0.303	−0.009	0.160	0.064	0.065	–	–	–	–	–	–
V	0.000	−0.206	0.181	−0.180	0.156	**0.637**	0.071	–	–	–	–	–
bVI	0.000	0.296	−0.049	**0.482**	−0.064	0.121	0.332	−0.060	–	–	–	–
VI	0.000	0.026	**0.483**	−0.170	**0.491**	0.151	0.049	0.249	−0.004	–	–	–
bVII	0.000	0.040	0.147	**0.367**	0.035	0.098	0.046	−0.040	**0.479**	0.096	–	–
VII	0.000	0.188	0.302	0.074	0.356	0.087	0.238	0.095	0.093	0.177	0.113	–

table (though the top half of the table is redundant); this is shown in Table 9. (As usual, we first calculated correlations in TdC's analyses and DT's analyses separately; Table 9 shows the average of the two.) A correlation is positive if two chords tend to co-occur, negative if they do not (it is always between −1.0 and 1.0). All correlations above .35 are shown in boldface; these correlations are also represented in Figure 4. Some interesting patterns can be seen here. bVII, bIII and bVI emerge as a group in which all three pairs are highly correlated; clearly, these three chords tend to co-occur. II, VI and III form another such group. These correlations offer evidence as to some kind of modal organisation, corresponding roughly to common-practice major and natural minor (bVII is not common in true common-practice minor). These correlations may also arise, however, simply out of local preferences for root motion by fourths (as seen in Tables 5 and 6), since the chords in both groups occupy consecutive locations on a line of fifths. The only other high correlation is between V and IV; no doubt this is because both of these chords occur in nearly all songs (see Table 2) and the songs they do not occur in are largely the same: rap and alternative-rock songs such as 'California Love' and 'Loser', and a few others such as 'My Generation' and 'London Calling'. (I occurs in all 99 songs; thus it has a correlation of zero with all other chords.)

We have focused here on root labels, since, as noted earlier, we find that more specific labels (e.g. chord quality and extension) are often highly ambiguous in rock. While we included such 'subcategorical' information in our original analyses, the chord-list format does not fully capture this information, since two adjacent chords of the same root and key are collapsed into one and assigned the label of the first chord; in cases where the two chords differ in quality or inversion (e.g. I going to I6), the label of the second chord will be lost. However, the chord list format still provides some subcategorical information and it is worth considering it briefly. With regard to chord quality, we can classify chords as major, minor, diminished or augmented; here we include seventh chords in the category of the triad on which they are built (for example, the dominant seventh is built on a major triad). By this measure, 75.8 per cent of chords in the corpus are major, 23.4 per cent are minor, 0.7 per cent are diminished and 0.1 per cent are augmented. The balance between major and minor chords is similar to that found in common-practice music, but diminished chords are much less frequent in rock (see Temperley 2009). With regard to inversion, 94.1 per cent of chords are in root position; here the contrast with

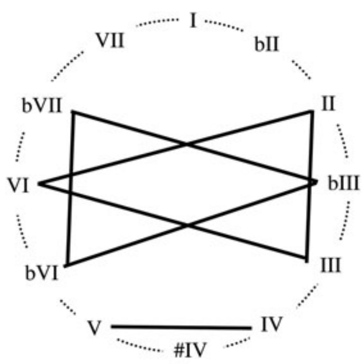

Figure 4. Lines indicate chord pairs with high correlations (above 0.35) in Table 9.

https://doi.org/10.1017/S026114301000067X Published online by Cambridge University Press

common-practice music is much more marked, where only about 60 per cent of chords are in root position (Temperley 2009).

Discussion and conclusions

In what follows, we review what we consider to be the primary findings that emerge from the statistical analyses above. We also consider some potential criticisms of our approach and some possible directions for future work.

The distribution of relative roots in the corpus shows a clear hierarchical pattern. The much greater frequency of IV and V over other chords gives these chords a privileged position among non-tonic harmonies. IV is, however, noticeably more frequent than V, suggesting that it has a unique and primary status. After IV and V, VI and bVII form a clear secondary category; the next most common harmony (II) is only about half as common as these. With regard to patterns of harmonic progression, the strong asymmetries of root motion found in common-practice music are notably absent in rock. The frequencies of chords in post-tonic and pre-tonic position are, for the most part, similar to their overall frequencies. Some positional preferences do, however, emerge. Most significantly, the frequency of IV as a pre-tonic harmony is somewhat higher than its overall frequency, suggesting that it often functions to prepare the arrival of the tonic in some way. The trigram data in Table 7 sheds further light on this, showing IV to be a versatile pre-tonic chord that is commonly approached by several different harmonies; by contrast, pre-tonic V is primarily approached by IV.

Generally speaking, the frequency of root motions is inversely related to their circle-of-fifths distance. Again, a striking symmetry is seen between the ascending and descending forms of each interval, with motion by ascending and descending fourths being by far the most common. One possible reason for this preponderance of fourth relationships may relate to the use of pentatonic scale collections in rock music. As discussed in Everett 2004 (his tonal system number five), certain rock styles build harmonic sonorities almost exclusively on members of the minor pentatonic scale. Note additionally that the roots of the standard first-position 'open-string chord voicings' on the guitar also form a pentatonic collection (C, D, E, G, A). Because they are particularly easy to play, these open chord voicings may be preferred harmonic choices irrespective of the tonic of a particular song. Since the fourth is the most common interval class of the pentatonic collection, random movements between members of this collection would statistically generate more fourth motions than those of any other interval class. However, if the pentatonic scale were the sole determinant of root motions, we would expect to see a balance between whole step root motions and those of major/minor thirds, since the interval-class vector of the pentatonic collection has an equal number of major seconds and major/minor thirds. Because root motions by whole step significantly outweigh those of major and minor thirds combined, other factors appear to be in play.

Our data suggest that harmonies in rock are overwhelmingly in root position, and that major triads are more common than minor triads (the latter difference is comparable to that found in common-practice music). With regard to trends across the five decades, the most notable pattern that emerges is a shift between the 1950s, when harmonic structures are almost exclusively confined to I, IV and V, and later decades, when a broader distribution emerges that remains fairly stable between 1960 and 2000. As for patterns of harmonic co-occurrence (Table 9 and

https://doi.org/10.1017/S026114301000067X Published online by Cambridge University Press

Figure 4), the current data is too preliminary for firm conclusions to be drawn, though there are strong suggestions of some kind of modal organisation.

One could re-evaluate the views of rock harmony discussed in the opening section of this paper, those of Stephenson, Moore and Everett, in light of our empirical data. We have chosen not to do this, however. As noted earlier, we realise that our brief summaries do not fully represent these authors' views. With regard to some of the large issues that arise in their writings, in particular the issue of modality, the picture emerging from our data is too incomplete to be of much help. With regard to other issues, such as the adherence of rock harmony to common-practice norms, our data is certainly relevant, but the degree to which it supports one position or another is, to some degree, a matter of interpretation. We do believe, however, that our data offers an interesting counterpoint to these authors' arguments. And we hope that it will prove useful to future studies of rock harmony, if only as a source of corroboration for more intuitive and informal observations.

We see various ways that the current project could be continued and extended. One obvious way would be to gather more data, by analysing more songs either from the *Rolling Stone* list or from other sources. We believe the current corpus allows fairly reliable answers to some basic questions, such as the overall distribution of chromatic roots. For other more specific questions, additional data would be desirable; for example, the identification of trigrams and larger patterns would certainly benefit from a larger sample. Further statistical analyses could also be conducted. One could, for example, examine the metrical placement of different harmonies. It has been suggested (Temperley 2001; Everett 2004) that rock reflects a strong preference for the placement of tonic harmony in metrically strong positions and that this is an important cue for tonal orientation; our data could be used to examine this claim. The issue of mode and palette, whose surface has been barely scratched here, could also be further investigated.

While the main purpose of our 'reduced' analyses was to provide data about rock harmony, they potentially serve another function as well. Our hierarchical method of analysis required us to partition each song into labelled sections (and sometimes sub- and super-sections). This results in a kind of formal analysis, which may be of interest in its own right. For example, the harmonic patterns that typify verses may be compared and contrasted to those patterns in bridge or chorus sections. The higher-level structures might also be of interest in themselves; one could examine the frequency of different formal patterns (such as different configurations of verse, chorus and bridge). As with harmonic analysis, of course, there is some disagreement between our formal analyses; this provides an indication of the subjectivity of form, which might be interesting to examine in itself.

It seems to us that the current project is open to criticism on at least two fronts. One concerns our method of analysing songs. Our analytical system treats harmonic structure as a completely 'flat', one-level sequence of harmonies. By contrast, many authors (Brown 1997; Burns 2008; Everett 2008b) have preferred to analyse rock harmony in a hierarchical (e.g. Schenkerian) manner, with some harmonies identified as structural and others as elaborative (or perhaps as purely linear, non-harmonic elaborations). We have acknowledged that many of the differences between DT's and TdC's analyses concerned cases where one of us analysed something as a harmony and the other analysed it as a purely linear event, part of a larger harmony. If a more hierarchical analytical notation were allowed, both of these analyses could be captured in a single representation: an underlying structural harmony elaborated by another chord in the foreground. There are problems with such an approach as well, however. The high degree of subjectivity in how exactly reductive analysis should be done (in rock music or any other style, for that matter) might greatly decrease the level of agreement among analysts. It is also not clear how aggregate statistics, regarding chord transitions, for example, would be gathered from a hierarchical representation. While 'single-level' harmonic analysis is by no means perfect, it does capture at least part of what is going on in rock harmony, and it has the advantages of being relatively immune to differences of opinion and easily amenable to aggregate data analysis.

The other issue that might be raised concerns our taking the *Rolling Stone* list to represent 'rock'. As we have acknowledged, 'rock' is clearly a fuzzy and ill defined term, and we cannot expect to find unanimous agreement as to its boundaries. We ourselves were somewhat puzzled by the inclusion of some songs on the list – songs that we feel could be considered rock only by an extremely broad definition of the term, in which jazz ('Georgia on My Mind'), rap ('Nuthin' But a 'G' Thang'), country ('I Walk The Line'), reggae ('No Woman, No Cry') and R&B ('People Get Ready') are all included under the umbrella of rock. We should recall that the people polled for the list were instructed to choose the 'greatest songs of the rock and roll era'; in some cases, they may have been more concerned with choosing the greatest songs of the era rather than choosing only central or unproblematic specimens of rock. It might be argued that what our corpus represents is not a single unified style, but perhaps several styles, each of which may have a more consistent harmonic logic than is reflected by the data we have presented above (this brings to mind Everett's six 'tonal systems').

It would be interesting to create a new corpus, or to refine the current corpus, in such a way as to include only songs that are judged to be unambiguously within the style of rock. In so doing, one might also create other categories, thus allowing a comparison of harmonic practice across categories. (We could of course make these judgements ourselves, but the current project is already heavily influenced by our subjective opinions in the analyses themselves, and we are reluctant to inject our opinions into the selection and categorisation of songs as well.) Another possibility would be, once again, to let the data speak for itself. To some extent, the songs of our corpus may naturally group into categories or 'clusters' by virtue of their harmonic palettes and patterns. (Our preliminary exploration of chord vectors, suggesting one group of songs that uses bVII, bIII and bVI, and another that uses II, VI and III, points in this direction.) Statistical clustering methods could be used to investigate this. The resulting clusters might prove to correspond quite well to chronological divisions and conventional stylistic categories (early rock 'n' roll, heavy metal, and so on), or perhaps they would suggest new classification schemes not yet considered.

Endnotes

1. Traditional common-practice theory more commonly speaks of descending *fifths* rather than ascending fourths. For the present purposes, however, it seems most logical to label each root motion with its smallest possible interval; thus we will use 'fourths' rather than 'fifths' throughout the article.
2. See, for example, the sources cited in the first paragraph of this paper; see also Larkin *et al.* (2003) and Miller (1980).
3. The '2008 Consumer Profile', as compiled by the RIAA (2008). In this study, the consumers were asked to classify their music purchases according to genre.

4. A number of other lists were also considered but found to be problematic for one reason or another. For example, Paul Williams's book, *Rock and Roll: the 100 Best Singles* (1993), is based on the opinions of only one person. Similarly, the 1999 VH1 list was neither as recent nor as explicit in its selection methodology as the 2004 *Rolling Stone* list. An excellent online resource that tracks such lists can be found at Franzon (2010).
5. One might question whether 'rock' is the same as 'rock 'n' roll'. Often the two terms seem to be used synonymously, as in the *Rolling Stone* text just mentioned. Sometimes rock 'n' roll is used in a more specific sense to refer to the very early years of rock (the 1950s). Since less than 15 per cent of the songs in the list were recorded prior to 1960, however, clearly the majority of the *Rolling Stone* voters did not understand the term this way.
6. Our analyses, programmes, and other related materials are publicly available at http://theory.esm.rochester.edu/rock_corpus/.
7. This assumes that the two analysts are using the same basic pitch framework, i.e. they agree on the pitch of individual notes. In that case, it seems unlikely that two analysts would agree on chromatic relative root while disagreeing on key (for example, one analyst analysing a chord as V/C and another analysing it as V/G).
8. Also of interest is the sheer number of chords in the two analysis sets (using the chord list representation). DT's analyses contained 9,300 chords, while TdC's contained 9,352; this suggests that TdC favoured a very slightly faster harmonic rhythm, though the difference is less than 1 per cent.
9. Similar disagreements occurred in '1999', 'Enter Sandman', 'Good Vibrations', 'My Generation' and 'Paranoid Android.'
10. Where the statistics being averaged are integers, the average is rounded down to the nearest integer.

References

Allmusic. 2010. http://allmusic.com (accessed 9 April 2010)

Brown, M. 1997. '"Little Wing": a study in musical cognition', in *Understanding Rock: Essays in Musical Analysis*, ed. J. Covach and G. Boone (Oxford, Oxford University Press)

Budge, H. 1943. 'A study of chord frequencies based on the music of representative composers of the eighteenth and nineteenth centuries', PhD thesis (New York, Columbia University)

Burns, L. 2008. 'Analytic methodologies for rock music: harmonic and voice-leading strategies in Tori Amos's "Crucify"', in *Expression in Pop-Rock Music: Critical and Analytical Essays*, 2nd edn, ed. W. Everett (London, Routledge)

Covach, J., and Boone, G. (eds.) 1997. *Understanding Rock: Essays in Musical Analysis* (Oxford, Oxford University Press)

Everett, W. 2004. 'Making sense of rock's tonal systems', *Music Theory Online*, 10

Everett, W. (ed.) 2008a. *Expression in Pop-Rock Music: Critical and Analytical Essays*, 2nd edn (London, Routledge)

Everett, W. 2008b. 'Pitch down the middle', in *Expression in Pop-Rock Music: Critical and Analytical Essays*, 2nd edn, ed. W. Everett (London, Routledge)

Franzon, H. 2010. *Acclaimed Music: The Most Recommended Albums and Songs of All Time.* http://acclaimedmusic.net (accessed 9 April 2010)

Holm-Hudson, K. (ed.) 2002. *Progressive Rock Reconsidered* (London, Routledge)

Huron, D. 2006. *Sweet Anticipation: Music and the Psychology of Expectation* (Cambridge, MIT Press)

Larkin, C. *et al*. (eds.) 2003. *The Billboard Illustrated Encyclopedia of Rock*, rev. edn (New York, Billboard Books)

Middleton, R. 1990. *Studying Popular Music* (Buckingham, Open University Press)

Miller, J. (ed.) 1980. *The Rolling Stone Illustrated History of Rock & Roll*, rev. edn (New York, Random House)

Moore, A. 1992. 'Patterns of harmony', *Popular Music*, 11, pp. 73–106

Moore, A. 1995. 'The so-called "flattened seventh" in rock', *Popular Music*, 14, pp. 185–201

Moore, A. 2001. *Rock: The Primary Text: Developing a Musicology of Rock*, 2nd edn (Aldershot, Ashgate Publishing)

Recording Industry Association of America. 2008. '2008 consumer profile'. www.RIAA.com, http://76.74.24.142/8EF388DA-8FD3-7A4E-C208-CDF1ADE8B179.pdf (accessed 9 April 2010)

Rolling Stone. 2004. 'The 500 greatest songs of all time'. *Rolling Stone*, 963 (9 December), pp. 65–165

Schoenberg, A. 1969. *Structural Functions of Harmony*, rev. edn (New York, W.W. Norton)

Stephenson, K. 2002. *What to Listen for in Rock: A Stylistic Analysis* (New Haven, Yale University Press)

Tagg, P. 1982. 'Analysing popular music: theory, method and practice', *Popular Music*, 2, pp. 37–67

Temperley, D. 2001. *The Cognition of Basic Musical Structures* (Cambridge, MIT Press)

Temperley, D. 2007a. 'The melodic–harmonic "divorce" in rock', *Popular Music*, 26, pp. 323–41

Temperley, D. 2007b. *Music and Probability* (Cambridge, MIT Press)

Temperley, D. 2009. 'A statistical analysis of tonal harmony'. http://www.theory.esm.rochester.edu/temperley/kp-stats/ (accessed 9 April 2010)

VH1. 1999. *VH1's 100 Greatest Songs of Rock and Roll*. Film. Dylan McDermott, host

Williams, P. 1993. *Rock and Roll: the 100 Best Singles* (New York, Carroll & Graf Publishers)

https://doi.org/10.1017/S026114301000067X Published online by Cambridge University Press