MARCUS CASTRÉN

RECREL

A Similarity Measure for Set-Classes

STUDIA MUSICA

4

SIBELIUS ACADEMY

HELSINKI 1994

Marcus Castrén
Recrel : A Similarity Measure for Set-Classes

ISBN 978-952-329-203-1 (PDF)
http://urn.fi/URN:ISBN978-952-329-203-1

MARCUS CASTRÉN

RECREL

A Similarity Measure for Set-Classes

By due permission of the Committee for Study and Research of the Sibelius Academy, Helsinki, to be publicly discussed in the Auditorium of the Sibelius Academy, Töölönkatu 28, 5th floor, on Wednesday, December 14th, at 12 o' clock noon, in candidacy for the degree of Doctor of Music

STUDIA MUSICA
4

SIBELIUS ACADEMY
HELSINKI 1994

MARCUS CASTRÉN

RECREL

A Similarity Measure for Set-Classes

STUDIA MUSICA

4

SIBELIUS ACADEMY

HELSINKI 1994

© Marcus Castrén and Sibelius Academy 1994

ISBN 952-9658-32-X
ISSN 0788-3757

YMPÄRISTÖMERKKI
MILJÖMÄRKT
411 011
Kierrätykseen sopiva tuote
Alhaiset päästöt valmistuksessa
Hakapaino Oy, Helsinki 1994

SIBELIUS ACADEMY
Department of Composition and Music Theory
Studia Musica No. 4, 1994.

Castrén, Marcus
RECREL: A Similarity Measure for Set-Classes
186 + 25 pages

■ ABSTRACT

In the field of pitch-class set theory, a number of different similarity relations have been developed with the purpose of identifying aspects of similarity between pcsets and set-classes. Out of these, the present study concentrates on similarity relations which compare interval-class or subset-class vectors. The focus is especially on similarity measures, i.e. similarity relations assigning a degree of similarity for two set-classes. Such a degree is given as a numeric value on some known scale of values.

Three main categories of similarity relations are identified. First, similarity relations not returning numeric values. Second, similarity measures comparing only two interval-class or subset-class vectors at a time. Third, total measures comparing all subset-class vectors belonging to two set-classes. This ordering of the categories is considered to reflect increasing descriptive powers. A set of relevance criteria which each similarity relation should fulfil is determined. 21 similarity relations presented previously in the pitch-class set-theoretical literature are then evaluated with these criteria. Other means of demonstrating the strengths and weaknesses of individual similarity relations are also used.

The RECREL similarity measure, belonging to the total measure category, is introduced and evaluated with the set of criteria. The values it produces are examined from different viewpoints. Also, it is used to analyse aspects of Arnold Schönberg's Op. 11, No. 1.

A demonstrational computer program constitutes a part of the study. With it the user can examine the various stages of a RECREL comparison, as well as manipulate RECREL values in a number of different ways.

■ ACKNOWLEDGEMENTS

I wish to express my gratitude to the many people who have helped me during the preparation of this thesis. Prof. Gary Wittlich (Indiana University) followed the project closely from its inception to its end. His highly professional advice was a combination of informed comments and suggestions, pertinent criticism and warm encouragement. Prof. Ilkka Oramo (Sibelius Academy) provided experienced advice on how to organize a multitude of hypotheses, outlines, subprojects, computer printouts, etc., into a finished study. Mr. Arto Wikla (Helsinki University) and Mr. Erkki Kurenniemi (HEUREKA Science Center) pointed to many an equation, formal expression or concept in the manuscript, and asked "Why?" Their illuminative explanations and suggestions made it possible to always arrive at a "Why not!" Prof. David Neumeyer (Indiana University) read the manuscript and made many insightful comments.

Dr. Andrew Bentley (Fazer Music Inc.) did not spare time or trouble taking me through the niceties of the English language. Mr. Mikael Laurson (Sibelius Academy) was an inexhaustable source of advice, helpful suggestions and support during the preparation of the computer programs essential for the realization of this study. Amanuensis Anna Krohn (Sibelius Academy) looked after many practical details in the last stages of the project. She never failed to cheer me up when saddled with everyday misfortunes, be it a broken printer or lost mail. Mr. Kimmo Salmio and Mr. Kalervo Koskimies (Sibelius Academy) lent their expertise in arranging the installation of the RECREL demo program on the Internet.

Also, I am grateful to the following institutions and foundations for their financial support: the Academy of Finland; the Sibelius Academy; Ministry of Education; Wihuri foundation; Kordelin foundation; Finnish Cultural foundation.

And finally, one who works with measures should especially appreciate things immeasurable: the support, encouragement and enthusiasm of family members, friends and colleagues. My dearest thanks to you all.

■ COPYRIGHT ACKNOWLEDGEMENT

Musical examples from Arnold Schönberg's
Piano Pieces, Opus 11, Number 1,
Copyright 1910 by Universal Edition.
Copyright renewed 1938 by Arnold Schönberg.
Used by permission of Universal Edition, Vienna.

■ TABLE OF CONTENTS

■ DEFINITIONS

Pitch-class (pc). The set of all pitches one or more octaves apart. The pcs will be labeled with numbers: C = 0, C sharp = 1,...B = 11.

Pitch-class set (pcset). A collection of pcs without duplication. The order between the elements is not determined. A pcset is written in curly brackets, the elements being separated by commas.

Transposition (T_n). A transformation adding some transposition interval n to every pc x in a pcset. Each sum (x+n) is taken mod 12.

Inversion (I). A transformation replacing every pc x in a pcset with its *inverse* (12-x). Each pcset T in set-class X has its inversion I(T) in the *inversionally related* set-class I(X). Of the inversionally related SCs, also shorter expressions, such as "I-related SCs," "I-pairs," etc., will be used.

Complement. The complement T_C of pcset T contains all pcs not included in T. Each pcset in set-class X has its complement in the *complement class* X_C.

Set-class (SC). A collection of pcsets mutually related by a transformation or a group of transformations. In this study, the transformations used to define a set-class are transposition and inversion. Two types of set-classes will be used. In the first of these, the pcsets are related by T_n only, in the other by T_n and/or I. We will use the terms T_n (transpositional) class and T_n/I class, respectively.

According to this definition, a set-class is simply a collection of pcsets. In the theoretical literature, however, SCs are routinely considered also as objects abstractly reflecting and representing the properties and characteristics of the individual pcsets they contain. SCs, then, could be thought of as entities having abstract elements, intervals and ics between the elements, ic contents, etc. We will follow this convention. In our discussion we will use almost exlusively set-classes.

When referring to the SCs, we will use the nomenclature given in Forte (1973a). Under T_n-classification the inversionally related classes will be distinguished by extra labels A or B. The class providing the "best normal order" (Ibid., 3-5, 11-13) is always the A class, its inversionally related class the B class. If neither A nor B appears in the name of a SC, it is inversionally symmetric.

Prime form. A pcset representing all member sets in a set-class. A prime form has no special musical importance among the member sets. The criteria by which it is determined, based on pc content, ordering and interval content, are mere conventions. In this study we will use prime forms as given in (Forte 1973a:179-81).

Cardinality. The number of elements in a pcset. Given pcset S of cardinality n, we will write #S = n. Set-class X of cardinality n refers to a SC with n pcs in each of its member sets. We will also use the term *n-pc* set-class, again meaning a SC whose member sets have cardinality n. Furthermore, when referring to a 2-pc SC, we can also use the term *dyad class*. Accordingly, a 3-pc SC is a *triad class*, a 4-pc SC a *tetrad class*, a 5-pc SC a *pentad class*, a 6-pc SC a *hexad class*, a 7-pc SC a *septad class*, an 8-pc SC an *octad class*, a 9-pc SC a *nonad class* and a 10-pc SC a *decad class*. All SCs of cardinality n constitute the *cardinality-class* n.

Intervals, Interval-classes (ics). Given successive pcs x and y, the *ordered pc-interval* between them equals (y-x) mod 12. If x and y are unordered, there are two (ordered) pc-intervals between them, (y-x) and (x-y). The pair of intervals, being complementary mod 12, forms an interval-class (ic). By convention each ic is represented by the smaller of the two intervals. The number of ics is seven, 0-6.

Interval-class vector (ICV). An array indicating how many instances of ics 1-6 can be found in a given pcset. An ICV can be given also for a SC. An ICV is written in square brackets. The ICV of the set-class X is denoted ICV(X).

Z-relation. SCs X and Y are Z-related if X ≠ Y and ICV(X) = ICV(Y).[1]

[1] Forte (1973a:21).

Subsets, Subset-classes, Inclusion. pcset S is included in pcset T (S is a subset of T) if each element in S is also an element in T. Set-class Y is (abstractly) included in set-class X (Y is a subset-class of X) if for every pcset S in Y there is at least one pcset T in X so that each element in S is also an element in T. The equation n!/(m! * (n-m)!) gives the number of subsets of cardinality m contained in a pcset of cardinality n. (m $\neq$ 0, m $\neq$ 1, m $\neq$ n).[2] A 10-element pcset, for example, has 252 5-element subsets. Likewise, the number of instances of pentad classes included in a decad class is 252.

Embedding function (Embedding number). Given SCs A and X, the embedding number of A in X, notated EMB(A,X), is the number of pcsets in A which are subsets of a given pcset in X. For example EMB(3-1,4-1) = 2, as any member set in 4-1 has two subsets being members in 3-1. Given the prime form of 4-1, {0,1,2,3}, the subsets belonging to 3-1 are {0,1,2} and {1,2,3}. The value of EMB(A,X) can be higher than zero only if #A $\leq$ #X and A is a subset-class of X.[3]

n-class vector (nCV), **Subset-class vector.** The n-class vector of set-class X, #X $\geq$ n, is an array of numbers comprising each of the values EMB(A,X), the argument A running through all SCs in the cardinality-class n in the order given by the Forte nomenclature.[4] The vector will be notated nCV(X). 2CV(X) is identical to ICV(X). 2CV(X), 3CV(X), ...#XCV(X) are together the subset-class vectors of X. The examples below give two versions of the 3-class vector of SC 5-15. The first one is being compiled under T_n/I-classification, the second one under T_n-classification.

```
3CV(5-15)  =  [1 0 0 2 2 0 0 4 1  0  0  0]
               1 2 3 4 5 6 7 8 9 10 11 12
```

The numbers between the square brackets are called *components*. The numbers below these are *indexes*, referring to the ordinal numbers in the SC names of the 3-pc classes in the Forte nomenclature. For example in the vector above, the component four above the index 8 indicates that four instances of the SC 3-8 are abstractly included in 5-15. The vector below gives some indexes twice, the left one always referring to the A class, the right one to the B class. Now the component in the index referring to 3-8A is two, as is also that in the index referring to 3-8B. The indexes will not always be shown.

---

[2] Forte (1973a:27).

[3] Lewin (1977), Morris (1987:89-90).

[4] Lewin suggests the name *M-class vector* for a vector like this in (1987:106-7). We use n in order to be consistent with the use of the symbol in the names of some similarity measures, like $MEMB_n$ and $\%REL_n$. Thus, $MEMB_n$ compares n-class vectors.

3CV(5-15) =	[1	0	0	0	0	1	1	1	1	0	0	0	2	2	1	0	0	0	0]
	1	2	2	3	3	4	4	5	5	6	7	7	8	8	9	10	11	11	12

%-vector. A normalised n-class vector. Instead of indicating how many instances of each SC of cardinality n is embedded in a given SC X, the vector gives the *percentual share* each n-class has of X's subset-class contents of cardinality n. To transform nCV(X) into the corresponding *n-class %-vector* of X, or nC%V(X), each component is divided by the sum of all components and multiplied by 100. The sum of the components is always 100. A component may be an integer or a fraction. For the sake of convenience, all components in the %-vectors will be rounded to the nearest integer. All actual calculations using the %-vectors, however, will be done to full accuracy. When transformed into 3C%V(5-15), the n-class vector 3CV(5-15) given above is as follows:

3C%V(5-15) =	[10	0	0	0	0	10	10	10	10	0	0	0	20	20	10	0	0	0	0]
	1	2	2	3	3	4	4	5	5	6	7	7	8	8	9	10	11	11	12

SC 3-1 is included in 5-15 proportionate to 10% of the total, SC 3-8A is included proportionate to 20% of the total, etc.

Similarity relation. A general term comprising a number of different approaches aimed at assisting in determining whether two or more pcsets or set-classes can be deemed similar or dissimilar to one another. Notions on which these relations are based include pc contents, ic and subset-class contents, SC family membership, etc. In the pcset-theoretical literature there is no universal agreement on how a similarity relation should be defined or what criteria a given theoretical construct should fulfil in order to be called one. In this study we will concentrate on *vector-based* similarity relations: context-free, non-transitive comparison procedures comparing the interval-class or subset-class vectors of two pcsets or set-classes at a time.

Similarity measure, Total measure. In this study, a similarity measure (a measure) is defined to be a vector-based similarity relation that compares the interval-class or subset-class contents of two pcsets or SCs at a time and produces a degree of similarity as a result. The degree is given as a numeric value on some known scale of values. A measure comparing subset-classes of all cardinalities mutually embeddable in two SCs will be called a total measure.

%RELn. A similarity measure used both as an independent measure and as a part of RECREL, usually to be evoked many times during a single RECREL comparison.

The subscript n indicates that any n-class %-vectors can be compared. In $\%REL_3$, for example, two 3-class %-vectors are compared. A $\%REL_n$ comparison is performed by taking the absolute values of the differences between corresponding nC%V components, adding these together and dividing the sum by two. The result, giving the extent to which the proportionate n-class distributions of the two SCs differ, lies always between 0 and 100, inclusive. The former indicates maximal similarity, the latter maximal dissimilarity.

k measure (k number). A similarity measure counting numbers of ic instances embedded mutually in two SCs. Given the interval-class vectors of two set-classes, the k value is the sum of the smaller components in each pair of corresponding ICV components.

Comparison group. The comparison group #n/#m contains all SC pairs {X,Y} such that X belongs to the cardinality-class n and Y belongs to the cardinality-class m. When n = m, a SC is not compared to itself, and a given pair {X,Y} = {Y,X} is counted for only once. When n ≠ m, the comparison group #n/#m contains $a_n{*}a_m$ SC pairs, a_n and a_m being the numbers of SCs in cardinality-classes n and m, respectively. When n = m, the number of SC pairs is $((a_n{*}a_n)-a_n)/2$. Under T_n/I-classification the number of SC pairs in the comparison group #3/#3 is 66, in #3/#4, 348, in #5/#6, 1900, etc. When discussing aspects common to all SC pairs in the comparison group #n/#m, we will speak of "#n/#m pairs," "#n/#m comparisons," etc.

Comparison groups can also involve ranges of cardinality-classes. The comparison group #3/#2-#12, for example, contains all SC pairs {X,Y} where X is a triad class and Y runs through the SCs of all cardinalities from 2 to 12. Again, a SC is not compared to itself, and when #Y = 3, a given pair {X,Y} = {Y,X} is counted only once. The comparison group #2-#12/#2-#12 contains all SC pairs except those involving SCs 0-1 and 1-1.

The *individual comparison group* X/#n contains all SC pairs {X,Y} where X is a constant referential SC and Y runs through the SCs in the cardinality class n. Correspondingly, the individual comparison group X/#n-#m, n < m, contains all SC pairs {X,Y} where X is the referential SC and Y runs through the SCs of cardinalities n to m. Whenever it occurs, the pair {X,X} is omitted.

Value group, The value group #n/#m contains the values that a given similarity measure returns to the SC pairs in the comparison group #n/#m. The measure providing the values can be selected freely.

■ CHAPTER 1
INTRODUCTION

Observing aspects of similarity between musical objects, as well as guiding one's attention with the help of these observations, are notions present everywhere in the work of an analyst or a composer. Being ever-present does not correlate with being familiar, however. The nature and dynamics of similarity assessments in music are questions of extraordinary complexity.

The advent of pitch-class set theory during recent decades has produced a wide range of studies concentrating specifically on similarity assessments.[1] This, no doubt, has to do with the fact that pcset theory offers some kind of "laboratory conditions" for a number of tasks. Each and every pitch combination in the tempered scale has only one pcset identity, each and every pcset only one set-class identity. A set-classification is an exact framework offering a wide range of possibilities for relating the elements it contains. We can identify properties of an individual SC, compare the properties of different SCs, and, whenever necessary, gather results from all comparisons in the entire SC universe.

Pitch-class set-theoretical tools developed with the purpose of assessing similarity between pcsets and set-classes are usually grouped under the heading *similarity relations*. In this study we will examine a number of these tools, as they constitute the theoretical background from which our main topic, the RECREL similarity measure, arises.

---

[1] The reader unfamiliar with the background, basic concepts and objectives of pitch-class set theory is directed to a source giving a general discussion on these questions. Among these are Straus (1990), Forte (1973a), Rahn (1980), Forte (1985) and Morris (1987).

In the pcset-theoretical literature, there is no universal convention or agreement on how a similarity relation should be defined, or what criteria a given theoretical construct should fulfil in order to be called one. Simply, the term comprises a number of different approaches aimed at assisting in determining whether two or more pcsets or set-classes can be deemed similar or dissimilar to one another.[2] Notions on which these relations are based include pc contents, ic and subset-class contents, SC family membership, etc.

Giving a general view of similarity relations is complicated, due to borderline cases, different interpretations of a single concept, etc. For example, a relation may be given without any concrete applications in mind, only in order to chart available possibilities.[3] Or, a relation may not be specifically associated with *similarity* considerations at all, but may be so consistently enjoyed by SCs with similar features that it is simply interpreted as a similarity relation, or at least as an important element in one.[4] A relation may even be expressly defined as *not* being a similarity relation, but may be used under circumstances where it assumes properties of one. In (1979-80:483) Rahn makes a clear distinction between an equivalence relation (transitive) and a similarity relation (non-transitive), but notes that there can be interaction between the categories: a recursively generated chain of pcsets forms an equivalence class which also can be viewed as quantifying similarity within itself by "nearness" in the chain.

Furthermore, a single concept or closely related concepts may be examined from altogether different viewpoints. According to Lewin (1977:194), for example, the interval-class vector (ICV) of Forte (1973a), the interval function of Lewin (1959) and the common-note function of Regener (1974) appear to describe the same musical phenomenon, but in fact model very different ones. The ICV is a rigorously harmonic concept, whereas the interval function is as rigorously contrapuntal in conception (Lewin 1977:201). The common-note function, in turn, tabulates relations not involving intervals as such, but rather transformations (Ibid., 218). In the most typical case of different interpretations, writers see

---

2 Many theorists exemplify their similarity discussions with the help of pcsets, whereas others use set-classes as their preferred objects of reference. Throughout this study, we will use set-classes (SCs) as our norm. When examining concepts and definitions presented elsewhere, we may modify them to suit this convention.

3 Rahn, for example, briefly outlines a similarity relation based on degrees of symmetry. He does not provide any analysis on its usefulness or descriptive powers (Rahn 1979-80:484).

4 Consider, for example, the central position of the complement relation in Forte's set complexes (Forte 1973a:93-100). In (Ibid., 78) Forte even states that it seems reasonable to regard the complement of a SC as a reduced or enlarged replica of that SC.

ICVs as describing either interval-classes (as distances between points), or instances of the dyad classes (as "2-note chord types").[5]

1.1 THE SIMILARITY RELATIONS TO BE EXAMINED IN THIS STUDY

It is not our intention to discuss the field of similarity relations in its entirety. This subject would merit a detailed study of its own. We will concentrate only on certain types of relations, i.e., those that are based on principles similar to those prevailing in our main topic, the RECREL similarity measure. To be more specific, we will examine *vector-based* constructs: context-free, non-transitive comparison procedures comparing the interval-class or subset-class vectors of two pcsets or set-classes at a time. We believe these comprise the most fruitful methods in assessing SC similarity yet offered. Within this category we will especially concentrate on *similarity measures*. We define them to be vector-based similarity relations comparing the interval-class or subset-class contents of two SCs at a time and producing a *degree* of similarity as a result. This degree is given as a numeric value on some known scale of values.[6] A similarity measure comparing subset-classes of all cardinalities mutually embeddable in two SCs will be called a *total measure*.

A number of well-known theoretical constructs will be excluded from our discussion because there do not seem to be commensurable criteria with which to analyse how their descriptive powers compare with those of the similarity measures. Among those excluded are the different systems of SC families that gather SCs into *Set Complexes* (Forte 1964 and 1973a, Kaplan 1991, described in Isaacson 1992), *Permutation Families* (Hämeenniemi 1983), *Regions* (Ericksson 1986), or *Harmonic Genera* (Forte 1988).[7] Some obvious notions distance our present interests from these, while conceding that ideally such a system might provide, in the words of Forte (1973a:93), a comprehensive model of relations among SCs. Firstly, some SCs X and Y may belong to a family of SCs not because they are suitably related to each other, but because they are suitably related to

---

5 In this study we will use both the interval-class and the dyad class viewpoints. The former will be used in the context of relations comparing only ic contents, the latter with relations comparing subset-classes of all cardinalities.

6 This definition is not a universal one. Forte (1973a:108) uses the term similarity measure when referring to his similarity relations that do not fulfil the criteria given here.

7 A comprehensive summary of most of these is given in Isaacson (1992:157-91).

some referential SC Z (such as the *Nexus SC)*.[8] In (1973a:108) Forte states that, when viewed from the standpoint of internal similarity relations, a set complex may be far from homogeneous. Secondly, the properties determining family membership may be such that the relation between some two members makes sense only when examined with respect to the family. In other words, the relation between the two is meaningful because of a specific context, not because of their internal properties.[9]

Other similarity relations which will be excluded from further analysis are, for example, the *R_p relation* (Forte 1973a:47),[10] a generalisation of this, the *Degree of Inclusion* (Regener 1974:207), and modifications of the latter, the *Degree of Degree of Inclusion* and its "adjusted" version (Rahn 1979-80:486-487). The *Exclusion Relation*, given in Clough (1983), gathers "families" of excluded SCs and is also beyond our present scope. Some concepts, like the above-mentioned *Interval Function* and *Common-note Function*, given in Lewin (1959) and Regener (1974), respectively, actually do involve vectors. In the service of a similarity relation, however, these would resemble ICVs to the extent that there is no need to examine them separately.

Concepts that also give cause for considerations of similarity but do not belong to the range of our present interests are those that in one way or another involve ordering. Among such concepts are, for example, the *Basic Interval Patterns* (Forte 1973a:63-73 and 1973b), the *Voice Pair Interval Sets* (Chapman 1981) and the *Constellations* of Hoover (1984). Finally, we will exclude similarity relations that compare superset-class vectors. We believe that the superset-class relations of a SC offer much weaker points of reference to similarity assessments than do its subset-class relations. Consider, for example, the 6-pc and 5-pc whole-tone classes 6-35 and 5-33, having intuitively an exceptionally high degree of similarity between them. *Every* subset-class of cardinality 5 or smaller in 6-35 is also a subset-class in 5-33. Moreover, the proportionate subset-class distributions

---

8 Morris notes in (1987:330 *n* 61) that the K and Kh relations offer an example of a similarity relation. Here, however, our focus is on all internal relations between the family members, not only those involving the referential SC with which a family is gathered.

For a definition and discussion on the nexus set concept, see Forte (1973a:101, 210).

9 Ericksson's Region 1, for example, contains both SCs 2-1 and 6-Z10, their respective ICVs being [1 0 0 0 0 0] and [3 3 3 3 2 1] (Ericksson 1986:102-3). It is difficult to see how the two ic contents could be deemed similar.

10 R_p relation holds between two SCs X and Y of cardinality n if at least one SC of cardinality n-1 is included in both of them. If two pcsets S and T share less than n-1 pcs but are member sets in two SCs enjoying the R_p relation, we say that the R_p relation holds between S and T but is weakly represented.

of the two classes are identical.[11] Their superset-class relations, by contrast, are entirely different. For example, all 7-pc superset-classes of 6-35 are instances of SC 7-33, whereas those of 5-33 are instances of 16 different T_n-classes.[12]

1.2 SIMILARITY RELATIONS IN PCSET-THEORETICAL LITERATURE

In the pcset-theoretical literature, the overall reception of similarity relations could be characterized as positive but usually not enthusiastic. Commentators, while perhaps criticizing a given relation or relations, recognize in principle that a trustworthy comparison method would be of interest and benefit. The conceptual basis of the relations - the possibility of identifying similarity between such abstract entities as pcsets and SCs - is not contested, but opinions vary on how concrete conclusions can be drawn from the results.

Hoover, being one of the sceptics, states that the nature of even the relatively low level of abstraction involved in a pcset is such that precisely defined relationships between pcsets cannot assure consistent musical relationships. A pcset relationship is hardly a complete determinant of the musical sense of all instances of the two pcsets (1984:165-166). Chapman, in turn, deems the roles of at least some similarity relations secondary at best. According to him, Forte's widely discussed R_0, R_1 and R_2 relations are abstract reflections of intervallic relations. They play only a minor role in most atonal analysis, usually appearing as supplemental observations upon analyses based on pitch-class set recurrence (1981:276-8). An interesting view is given by Beach. He recognizes that different degrees of SC similarity exist, but doubts the usefulness of similarity measures. Identifying just minimal and maximal similarity is enough, as "enumeration of other levels would only serve to encumber an already complex and sophisticated theory of sets and set relations" (Beach 1979:11, quoted in Isaacson 1992:19-20).

Rahn, being more optimistic, sees similarity relations in a dynamic context, as integral parts in a hypothetical "general theory of harmony." Such a theory needs a context-free relation of similitude or reasonable facsimile for its basis (1989:9). A "theory of instances," in turn, is needed as a front end to choose the

---

[11] A distribution like this indicates how large a *percentual share* each embedded class has among all subset-classes of the same cardinality. This concept is the basis of several similarity measures, including RECREL.

[12] During the initial stages of the RECREL project, a version comparing also superset-class vectors was tested. It was abandoned for the reasons just stated.

A number of similarity relations involving superset-classes are presented and discussed in Isaacson (1992:136-56).

individual member sets. The latter theory could be based on common-pc considerations and a criterion of least pc distance between moving (noncommon) pcs (1989:5). Some writers, when arguing for their relations, do not point only to conceptual or intuitive notions, but also straight to listening experiences. Morris, for example, states that his SIM relation provides a rationale for the selection of SCs that ensure predictable degrees of aural similitude (1979-80:446). Later on he notes that the value zero (indicating maximal similarity in SIM) does not necessarily mean that the SCs involved are equivalent under T_n and/or T_n/I as they can be also Z-related. According to him, if two Z-related SCs are comparably represented in a musical setting, they will have a good deal of sonic similarity (Ibid., 447).

According to the most optimistic opinions, a similarity relation can aid greatly in our understanding of both conventional and modern music (Solomon 1982:104-106, on the R relation), and be an important and useful tool for music analysis (Isaacson 1990:25, on the IcVSIM relation).

1.3 QUANTITATIVE AND QUALITATIVE SIMILARITY

Pcset-theory does not assign any qualitative characteristics to any pcset-theoretical objects. Consequently, all assessments of SC similarity are purely quantitative.[13] We observe, for example, the extent to which two subset-class contents consist of corresponding elements, and equate high extents with high degrees of similarity. No other aspects, measurable or non-measurable, effect the outcome.

This principle, determining a precise "testing ground" for the measurements and defining similarity only with respect to it, goes often unnoticed. The reason is that usually it is not in conflict with qualitative similarity assessments. An individual observer may associate SCs with strong qualitative characteristics and still agree that SCs with disjoint subset-class contents are dissimilar, those with near-identical ones similar. Borderline cases do exist, however. Suppose we have SC pairs {X,Y} and {Z,W} so that X shares exactly half of its interval-class instances with Y, Z half of its interval-class instances with W. If shared-instance extent is set to be the sole criterion of SC similarity, the pairs are equally similar and the non-shared instances have no effect whatsoever. The observer, however, might assign qualitative characteristics to the non-shared instances, for example so that those in both X and Y seem consonant, whereas those in Z seem

13 See related remarks in Alphonce (1974:153) and Isaacson (1992:79).

clearly different from those in W. The former pair, then, would be favored in terms of qualitative similarity.

The decision to restrict our observations to only quantitative aspects arises from cases like this. Introducing non-pcset-theoretical, non-measurable elements could have benefits when relating isolated cases, like pairs of small-cardinality SCs familiar from traditional harmony. But the effects these elements would have among all comparisons, or the possible distortions they would cause, are impossible to determine accurately. For example, if we deem ic1 closer to ic2 than to ic4, what is the actual weighting function which exposes the ic1-ic2 similarity during a measurement? If some function would work fine with triad class pairs, would it do so also with nonad class pairs, etc?

This is far from suggesting that qualitative assessments should be ignored. They could, and eventually probably will, have interesting interaction with assessments of SC similarity. But it is important to understand that coordinating these aspects requires a step beyond the scope of pcset-theory in general and this study in particular. Therefore, in the following discussion, by the concept "similar" we will mean only similarity with respect to the aspects we observe.

1.4 EVALUATING THE VALIDITY OF A SIMILARITY RELATION

In principle, the analysis of a similarity relation is carried out in two stages. The first stage comprises the analysis of the comparison procedure itself, and is a relatively straightforward task: Is the relation generally well-conceived, or do some aspects in it appear to be artificial or counterintuitive?; Does it compare only the properties it is designed to compare, or do other, undeclared aspects also have an effect?; If so, are the results somehow distorted?; Is the way with which the results are presented meaningful, or is there an evident possibility of misinterpreting them?; etc.

The second stage is the analysis of the actual descriptive powers of the relation, a task with a sort of built-in paradox. I.e., when examining results that are produced by an exact theoretical construct comparing exactly defined objects (SCs) in a space of exactly known limits, the theorist is assisted by little else than intuition. The final evaluation is a combination of many individual assessments: Is the scale of values from a given comparison group credible when related to that from another?; If some SC pair X is experienced to be closely similar, pair Y even more so and pair Z nothing but, do the corresponding values seem to reflect this meaningfully?; Does intuition confirm or contradict results suggesting that all SCs of cardinality n are equally similar to all SCs of cardinality m?; etc.

Obviously then, conclusions drawn from absolute measurements are of relative nature. The exactness of the formal foundations of pcset-theory does not guarantee exactness for a similarity relation. Like any concepts involving a highly complex mixture of different considerations, the relations are to be improved by subjecting them to careful analysis, criticism and suggested modifications. These help to bring about novel hypotheses and discard unsatisfactory ones, gradually increasing our awareness of the subject.

1.5 ON THE OBJECTIVES OF THIS STUDY

We believe that a reliable method for identifying SC similarity would have a wide range of important applications in analysis, composition and music theory. In our opinion, however, most of the previously presented similarity relations perform this task less than successfully.

Some of them do not distinguish between degrees of similarity at all, an aspect we think is in stark contrast with the very basis of assessing SC similarity. Our intuition does not suggest a sharp distinction between "similar" and "not similar," comparable to observing a light that can only be switched on or off. Obviously, a gradation allowing intermediate degrees between the extremes must be involved. Some other similarity relations do produce values indicating degrees of similarity, but are based on comparisons between limited reference materials, such as ic contents. This, we believe, weakens their descriptive powers. A well-known deficiency resulting from this is the inability to discriminate between Z-related SCs. Furthermore, some relations are of limited reliability because the methods with which they compare SCs contain counterintuitive features, even clear conceptual errors. What is claimed to be measured does not entirely coincide with what is actually being measured.

One of the main objectives of this study is to present this criticism in a concrete and detailed manner. The second objective, the principal one, is to offer an alternative to existing similarity relations. It is our belief that comparing entire subset-class contents constitutes the most reliable basis for assessing SC similarity. We will analyse the existing total measures, discussing ways to share their strengths and avoid their weaknesses. The conclusions we draw are then forged into concrete principles incorporated in the RECREL similarity measure.

1.6 THE CHAPTERS IN OUTLINE

In chapter 2, notions relevant to assessments of SC similarity are examined at an abstract level. The aim is to identify what sort of conditions a similarity relation must fulfil before it can be said to produce meaningful and reliable results. The conditions are presented as a set of criteria, with the help of which the validity of the various similarity relations offered in the literature can be evaluated.

The actual evaluations are given in chapter 3. Every relation is examined with respect to each criterion in the set of criteria. Strengths and weaknesses outside the scope of the criteria are also identified and analysed, usually with the help of individual SC comparisons. To make the evaluations as accessible as possible, all relations are presented in a strictly uniform manner. The entries are: a characterization of the relation with respect to a categorization given at the beginning of the chapter; a short verbal description of the comparison procedure; a mathematical formula (whenever the author of the relation provides one); an example or a few examples; a list of the evaluation criteria the relation fulfils; information about the entire set of values (for similarity measures only); information about individual value groups, given as a table (for measures only); analysis. The analysis sections of some relations are divided into further sections, containing special topics such as comparisons between different similarity relation categories, analysis of values belonging to inversionally related and Z-related SC pairs, etc.

Using the concepts introduced in chapters 2 and 3, the RECREL similarity measure is presented in chapter 4. Due to its complexity the measure will be introduced in three stages, which describe the comparison procedure in increasing detail. Examples are followed by two different RECREL formalisations, and finally, we examine which evaluation criteria RECREL fulfils.

The values which RECREL produces are analysed in chapter 5. The aspects examined first are those of the most general nature: the lowest and highest values, the number of distinct values, the distribution of the values, the values all SCs in a given cardinality-class produce with SCs of all cardinalities, the value groups #n/#m, etc. After this, a study is made of what sort of values RECREL produces to SCs that enjoy relations often associated with close similarity. There will be five categories: inversionally related SCs, Z-related SCs, complement pairs, SCs of cardinality n and their subset-classes of cardinality n-1, M-related SCs.

In chapter 6 RECREL is evaluated from another point of view, as an analytical tool. The work to be examined is Arnold Schönberg's piano piece Opus 11, Number 1. First, results concerning prominent SC materials and the use of those materials are adopted from two earlier analyses offered in the literature. An examination is then made of how the suggested harmonic characteristics of the music (resulting from specific types of SC arrangements) correlate with results generated with RECREL. Next, several palindromic SC successions are identified in the music and analysed with RECREL. Finally, the measure is used to define certain types of SC families. These, in turn, are used to analyse passages containing unusually high concentrations of closely related SCs.

A glossary summarizes the most important concepts associated with RECREL. The appendix contains a manual for a demonstrational computer program. With the program the user can examine the various stages of a RECREL comparison, as well as manipulate RECREL values in a number of different ways. Instructions in how to obtain a copy of the program are given in the appendix.

■ CHAPTER 2
SIMILARITY RELATION EVALUATION CRITERIA

2.1 INTRODUCTION

In this chapter we will examine what sort of conditions a similarity relation should meet before it can be said to serve its purpose well. First, in section 2.2, we analyse a previously presented set of criteria identifying some of these conditions. Then, in section 2.3, we introduce a set of criteria of our own. This criteria will be our reference point when we analyse various similarity relations in chapter 3. The criteria are six in number, one of them being further divided into four subcriteria. The criteria are connected with the cardinalities of the compared SCs, cardinalities of the subset-classes with which they are compared, the scales of values the measures produce, etc. Each criterion will be analysed in detail in sections 2.4.1 - 2.4.5.

At the end of the chapter we discuss two topics closely related to the criteria. The first one is the status of different subset-class cardinalities in the service of a similarity measure: Does it make a difference if a measure processes subset-classes of all cardinalities as one large group, or each cardinality as an independent entity?; Do subset-classes of different cardinalities seem to reflect similarity in a consistent manner?; etc. (Section 2.5). In section 2.6, we analyse the relation between a similarity relation and a set-classification.

2.2 ISAACSON'S CRITERIA

In the literature, authors often point to unsatisfactory features of previously presented similarity relations, and point to favourable ones in those they are about to introduce. It is rare to find detailed analysis of minimum conditions which a relation must meet in order to be valid, however.

Isaacson (1990:2) suggests three criteria for similarity relations comparing interval-class contents. A relation should (1) provide a distinct value for every pair of SCs, (2) be useful (not just usable) for SCs of any size, (3) provide a wide range of discrete values.

The first criterion rejects a number of relations for not being similarity *measures*. They provide an insufficient degree of discrimination, the yes-or-no type of outcome indicating only whether two SCs enjoy the relation or not. We assume that the second criterion means that a measure must produce meaningful results from comparisons between SCs of all cardinalities in order to be useful. What is more, this criterion discourages the use of some measures, as they are designed for SCs of the same cardinality only. The meaningfulness of criterion (3) is not as evident as that of the two previous ones. The set of values produced by some ideal similarity measure would truthfully describe the degrees of similarity between all SC pairs. It is difficult to see how we could meaningfully place any concrete expectations on this set of values in advance. It might consist of a wide range of discrete values, or it might not. Its properties would reflect the properties of the SCs, not those of the measure. We deem the third criterion more of a recommendation than a condition on a par with the two previous ones.

The three criteria provide a basis for the analysis of different similarity relations in Isaacson (1990). The focus is strictly on ic content similarity, to the extent that some of the relations are examined only with respect to ic contents, although they would also allow the whole subset-class contents to be compared (Ibid., 8-13). Our own set of criteria, to be introduced below, reflects our conviction that comparing the entire subset-class contents of two SCs is a better starting point for a similarity relation than comparing ICVs only.

2.3 SIMILARITY RELATION EVALUATION CRITERIA

A similarity relation should:

C1) allow comparisons between SCs of different cardinalities
C2) provide a distinct value for every pair of SCs
C3) provide a comprehensible scale of values, so that
- C3.1) all values are commensurable
- C3.2) the end points are not just some extreme values, but can be meaningfully associated with maximal similarity and dissimilarity
- C3.3) the values are integers or other easily managable numbers
- C3.4) the degree of discrimination is not too coarse or unrealistically fine

C4) produce a uniform value for all comparable cases
C5) observe mutually embeddable subset-classes of all meaningful cardinalities
C6) observe also the mutually embeddable subset-classes *not* in common between the SCs being compared

2.4 ANALYSING THE CRITERIA

2.4.1 Criteria C1 and C2

Our second criterion is the same as the first one in Isaacson (1990), while C1 is connected to Isaacson's second criterion.[1]

2.4.2 Criterion C3: A Comprehensible Scale of Values

The third criterion is divided into four parts, listed in order of decreasing importance. C3.1 is one of the most important in the whole set of criteria. C3.2, when met, is useful as it gives natural points of comparison when relating different values to each other. C3.3 is as much a recommendation as it is a requirement, stating simply that easily managable values make a similarity measure more convenient to use. C3.4 is of a very general nature, calling for scales of values

[1] As it may be difficult to remember precisely what each criteria stated, we shall at times refer to them with informal reminders, such as "the cardinality difference criterion C1," "the value commensurability criterion C3.1," "the total subset-class contents criterion C5," etc.

with degrees of discrimination which neither underestimate nor overestimate our intuitive ability to identify grades of similarity.

2.4.2.1 Criterion C3.1: Commensurable Values

C3.1 states that there must be only one uniform scale of values for all comparisons, regardless of the cardinalities of the SCs being compared. Let us analyse the meaning of this criterion with a concrete example. Suppose we measure ic content similarity between pairs of SCs by first adding together the corresponding ICV components if both are nonzero and then adding together the individual sums. In principle, the higher the final sum, the higher the degree of similarity between the classes.[2] Suppose we compare in this way all SC pairs in the three comparison groups #3/#3, #3/#4 and #4/#6. In the resulting value group #3/#3, all 66 values lie between 0 and 5, inclusive. The 348 values in value group #3/#4 lie between 0 and 9. In value group #4/#6, the 1,450 values lie between 5 and 21.[3]

Let us examine only instances of the value 5. In value group #3/#3 it is the maximum value. In #3/#4 it is situated in the middle range of values and in #4/#6 it is the lowest value. When relating the values to each other, we are presented with two alternatives. On the one hand, we could conclude that even if the numbers are the same, they are from three different value groups and represent three different degrees of similarity. On the other hand, we could conclude that because the pairs share a uniform value, they must represent a uniform degree of similarity. The argument in favour of the former alternative is obvious. The numbers of ic instances involved in a #3/#3 comparison, a #3/#4 comparison and a #4/#6 comparison differ considerably. A value being uniform in absolute terms is anything but uniform when related to the sizes of the ic contents involved. For two triad classes the sum of components in the two ICVs is 6. For a #3/#4 pair the corresponding figure is 9, for a #4/#6 pair, 21. Obviously, 5 out of 6 suggests stronger similarity than 5 out 9, or 5 out of 21.

If we choose the latter alternative and take each value at its "face value" instead of scaling it to the sizes of the ic contents, the measure would offer us the following results: the most similar #3/#3 pair represents the same degree of similarity as the most dissimilar #4/#6 pair; the most similar #4/#4 pair represents a lower degree of similarity than the most dissimilar #4/#8 pair; when compared

---

[2] The measure in question is Rahn's $MEMB_2$, to be examined in detail in chapter 3.

[3] The comparison groups were compiled under T_n/I-classification.

to both itself and to SC 12-1, every SC X of cardinality 11 or lower would get a higher value from the comparison $MEMB_2$(X,12-1) than from $MEMB_2$(X,X).[4]

Results like these would be, of course, nothing short of absurd, and show that these seemingly uniform values are in fact incommensurable. This same problem of different scales of values for different comparison groups applies also to a number of other measures besides $MEMB_2$ (s.i., SIM, k, *sf*, IcVSIM, TMEMB). Some writers recognize this problem (Morris 1979-80:450), others do not. It can be avoided either by avoiding comparisons between values from different value groups, or, more fruitfully, by developing a modification which eliminates the undesired property (ASIM and $IcVD_1$ from SIM, ak from k, ATMEMB from TMEMB, $\%REL_n$ from *sf*).[5] In our opinion, inability to meet the criterion C3.1 is a defect which seriously limits the usefulness of a similarity measure.

2.4.2.2 Criterion C3.2: Meaningful Extreme Values

It is beneficial if the numerical limits of a scale of values are some easily understandable numbers, such as 0 and 1 or 0 and 100.[6] Extreme values such as these provide a clear conceptual frame to which all intermediate values can be easily related. When this is not so, the maximum value cannot be intuited from any evident properties of the relation, it being rather some number representing the highest result obtained from all comparisons. This can mean that the conceptual frame is not evident and a given value must be processed further in one's mind to understand its position on the scale. It is easier to relate 0.75 to 1 or 75 to 100 than, say, 2.73 to 3.64, even though the first figure of each pair is three quarters of the latter in each case. 3.64 is the highest value produced by the IcVSIM relation.

2.4.2.3 Criterion C3.3. Easily Managable Values

It is beneficial if each value is an integer or some other type of easily managable

---

[4] The results are from the $MEMB_2$ value group information table (section 3.6.10) and from additional comparisons.

[5] All these measures will be analysed in detail in chapter 3.

[6] For some measures the value zero indicates maximal dissimilarity while increasing values indicate increasing similarity. For others the role of the values are exactly the opposite, zero indicating maximal similarity and increasing values increasing dissimilarity. The latter ones, as Rahn (1979-80:489) points out, are strictly speaking measures of *dissimilarity*. When analysing the measures we will not make this distinction separately, as it becomes evident from the context anyway.

number. When discussing the s.i. measure proposed in Teitelbaum (1965), Lord criticises it for offering results mostly in the form of irrational numbers, a convention which he sees as lacking clarity and accessibility (1981:111 *n* 7).[7] The values produced by the IcVSIM relation are almost always irrational numbers as well, a feature Isaacson admits will undoubtedly trouble some readers (1990:19).

2.4.2.4 Criterion C3.4: A Reasonable Degree of Discrimination

If the values produced by a similarity measure are not integers, the number of decimal places to which they are rounded must be decided at one point or another. This, in turn, has a straight effect upon its degree of discrimination, or, using the optical term used by Rahn (1979-80), resolution. Discussion of this topic does not occur in the literature. Writers either start by using a certain numbers of decimal places right away (Rahn 1979-80, discussing the ak measure; Morris 1979-80:450), or do not provide any examples and ignore the whole question (Rahn 1979-80, discussing ATMEMB; Lewin 1979-80b). A possible reason for this might be that the writers did not have sufficient computer capacity at their disposal to calculate all values, putting resolution concerns out of reach.

Two approaches seem to be naturally available for determining the accuracy to adopt. The first is the one adopted in Isaacson (1990) and (1992).[8] The values are calculated with an accuracy that is simply fine enough to distinguish between extremely close but still non-identical values, no matter how fine the gradation must be in order to achieve this.[9] This approach is valid, of course, as it in a sense reveals the "true" resolution of a measure. We disagree, however, with a conclusion Isaacson seems to have drawn from this, a conclusion to which he repeatedly refers but never clearly states: in order to be useful, a measure should provide a "wide range of discrete values," "produce fine distinctions between values" and "show fine gradations of similarity." (1990:2, 19, 22). A measure can provide only a handful of values and still be very good.

The other alternative, to be adopted in this study, is a more practical one. All non-integer values will be rounded in a uniform manner, to two decimal places. This means, for example, that if the extreme values produced by a mea-

[7] In fact Teitelbaum uses square root notation, a convention which clearly meets C3.3.

[8] Isaacson does not explain his strategy but it becomes clear from the context.

[9] When Isaacson illustrates an aspect of Lewin's REL measure, he gives a group of values rounded to eight decimal places, thus using a scale containing one hundred million grades (1992:164).

sure are 0 and 1, the total number of grades on the scale 0, 0.01, 0.02,...1 is 101. It may turn out, however, that many or perhaps even most of the grades do not appear in the final value scale, as they are not values of any actual SC pair comparisons. The scale can be considerably coarser than the gradation would allow.

When adopting the latter alternative, we in fact reduced the resolution some measures were capable of providing. We nevertheless believe that the present resolution is fine enough. When a similarity measure is tested, it is most important of all to identify potential inherent distortions and assess overall usefulness, not to show minute gradations of similarity that exceed the resolution of the very tool with which the evaluation is done, i.e., intuition.[10] Comparing intuitive degrees of similarity to measured degrees of similarity may strongly support the validity of a given measure, but such comparisons certainly do not encourage us to take the resolution to extremes.[11]

When we come to evaluate the measures in chapter 3, we will provide the number of distinct values which each one produces when examined at our adopted accuracy. The numbers may differ from those given in Isaacson (1992) due to the different accuracy and the fact that Isaacson examines value groups #2-#10/#2-#10 while we examine value groups #2-#12/#2-#12. In most cases the number of distinct values is above 30 and below 150. We will determine that all measures fulfilling the distinct value criterion C2 also fulfill the meaningful degree of discrimination criterion C3.4, with the sole exception of Rahn's TMEMB measure. The number of distinct values it produces is 877.

The distinct value count will not be considered as an important indicator of usefulness. On the contrary, we believe that if a measure is incapable of meeting the value commensurability criterion C3.1, as many measures are, the count is useless: one value can indicate many degrees of similarity, and one degree of similarity can be indicated by many values.[12] In these cases, only distinct value

---

10 It does not seem very bold to assume that if a measure divides the distance between maximal similarity and maximal dissimilarity into 8826 steps, its resolution surpasses our intuitive one. 8826 is the number of distinct values produced by Lewin's REL measure, as analysed by Isaacson in (1992:122).

11 Given, for example, a scale from 0 to 1 with increasing values indicating increasing dissimilarity, it is difficult to envisage circumstances under which we could draw meaningful analytical conclusions from the fact that SC pair X has the value 0.1001 and SC pair Y 0.1002. Or, we do not expect our intuition to confirm that SC pair A is two times as similar as SC pair B and three times as similar as SC pair C, their values being n, 2n and 3n, respectively, etc.

12 Comparing the distinct value counts in these cases can be outright misleading. The k and SIM measures, for example, are connected so that the k value of a SC pair can be calculated from its SIM value and vice versa. (Section 3.4.1.1). The two measures have identical numbers of distinct values in corresponding value groups. However, the total number of distinct values is 44 for SIM and 35 for k. A given k value *v*, indicating many degrees of similarity in various k value groups, can corre-

counts within individual value groups are of interest. Likewise, the average of the value group #2-#12/#2-#12 is of importance only if the measure in question meets criterion C3.1.

2.4.3 Criterion C4: A Uniform Value for Comparable Cases

This criterion states that a similarity measure should not be effected by SC properties other than those it professes to measure. If it adopts as its basis a certain aspect of similarity, it should produce the same value for all SC pairs whose type of similarity is uniform from the point of view of the chosen aspect.

Let the comparison group #3/#3 be our test material in examining this criterion. Many measures launch ICV comparisons by taking the differences between corresponding components. The basic principle is to associate increasingly similar numbers of corresponding ic instances with increasing degrees of SC similarity. The degrees are only expressed indirectly in a sense, since we record the number of ic instances *not* in common: smaller differences mean closer similarity. Other measures apply the same basic principle but approach it from the opposite direction, by recording the number of ic instances in common. For example, since all triad classes contain three ic instances (the sum of components in their ICVs is three), every SC pair in the comparison group #3/#3 falls into one of just three categories. A given pair may have (a) no ic instances in common (b) one ic instance in common (c) two ic instances in common.[13]

The SC pairs in each of these categories are examples of what we mean by "comparable cases" in the definition of criterion C4. They are comparable both with respect to their cardinalities and to the numbers of ic instances they have in common. Consequently, if they are compared with a similarity measure that is based on observing numbers of ic instances in common, no other aspect should stand in the way of them getting identical results from it. A number of similarity measures fail this rather obvious criterion, however. The methods used to calculate the values take into account, perhaps inadvertently, aspects other than ic contents as well. These measures require, at one stage or another, that differences between corresponding vector components be squared. They have the unfortunate tendency of treating vectors with similar-sized components differently from

---

spond not to just one, but to many SIM values.

[13] The result is valid under T_n/I-classification. The number of ic instances in common between two SCs is given as their *k number*. The concept is from Morris (1979-80:448) and will be examined in detail later.

those with highly different components. Not one of the theorists proposing a measure with this property has identified this inherent defect.

Let us suppose we have two pairs of SCs, {X,Y} and {Z,W} which are comparable in the sense described above. The ic instances in ICV(X) and ICV(Y) are evenly distributed. ICV(Z) and ICV(W), by contrast, have what we will call *peaked* ic distributions, the peaks being located in different indexes. When we take the differences between the corresponding components, ICV(X) and ICV(Y) produce small differences and relatively small squared differences. But when ICV(Z) and ICV(W) are compared, the peaks do not "neutralize" each other, resulting in large differences and even considerably larger squared differences. Whatever further processing we do to these two sets of squared differences, the disproportionately large elements in the latter cause the similarity measure to not produce uniform values for {X,Y} and {Z,W}.

Measures with this feature are meant to be based on the extents of shared instances in two ic contents, but they end up being based on another notion as well, i.e., whether the *non-shared* instances are evenly distributed or in "piles." From the point of view of ic content similarity the latter aspect is not only irrelevant but outright distorting. Sensitivity to the non-shared instance distribution introduces an artificial element to the comparison. The IcVSIM value group #3/#3, for example, contains 7 distinct values, despite the already mentioned fact that, from the point of view of numbers of ic instances in common, triad class pairs compared under T_n/I-classification can differ only in three ways.

Let us compare two pairs of SCs, {3-1,3-9} and {3-7,3-8}. The ICVs are given in Ex. 2.1.

EXAMPLE 2.1: Interval-class vectors of four 3-pc classes.

```
ICV(3-1) = [2 1 0 0 0 0]
ICV(3-9) = [0 1 0 0 2 0]
ICV(3-7) = [0 1 1 0 1 0]
ICV(3-8) = [0 1 0 1 0 1]
```

The classes in both pairs share one instance of ic2. It is difficult to see how one could claim that, from the point of view of ic contents, the {3-1,3-9} pair represents a higher degree of dissimilarity than the {3-7,3-8} pair does. Yet several similarity measures sensitive to the non-shared instance distribution suggest it does. The reason is that both ICV(3-1) and ICV(3-9) contain a "peak" in the form of a component 2, whereas all nonzero components in ICV(3-7) and ICV(3-8) are of value 1. It is possible to defend the peakedness-sensitive measures by saying that they utilize widely applied mathematical concepts, such as the standard deviation function in IcVSIM. We disagree with this view. Such a concept has no guaranteed musical validity and cannot be used as an excuse if a measure produces counterintuitive results.

2.4.4 Criterion C5: Observe Mutually Embeddable Subset-classes of All Meaningful Cardinalities

Here we call for the expansion of the subset-class materials participating in the comparisons. The expression "all meaningful cardinalities" is an intentionally ambiguous one, since participating subset-class cardinalities may be selected differently in different measures. For example, a measure may or may not take advantage of the fact that a SC is its own subset-class.

The reasons for comparing all mutually embeddable subset-classes and not just dyad classes (ics) are obvious. First, in a very general sense, the more points of reference we have between the SCs to be compared, the more accurately we can hope to demonstrate structural similarities or differences between them. As we identify some properties of a SC with the help of its ICV, it does not seem counterintuitive to assume that we could identify some of its other properties with its other vectors. Second, we want to discriminate between inversionally related and Z-related SCs. They are identical from the point of view of ic content-based similarity measures.

When observing some collection of elements together with its all subcollections, one could intuit that the largest subcollections preserve more of its substance and structure than small ones and are therefore of greater interest. There is, for example, an immediate sense of similarity between some 9-element chord and its nine 8-element subchords. A counterargument exists as well, however, which prevents us from establishing a simple correlation between the cardinality of a subcollection and its importance. It is articulated by Morris as he argues that intervals and ics are the backbone of our audition. In his opinion large shared subset-classes influence auditory comparisons of SCs as well, but it can be suspected that they might become too large and produce too fine a measure for us to hear (1979-80:458 *n* 14).[14] It seems he may have changed his mind to a certain ex-

[14] Despite the fact that Morris's own SIM relation compares ICVs only, he refers to entire subset-class contents when discussing similarity relations in general. For instance he criticises a relation for deeming SC pairs {6-1,6-32} and {6-1,6-2} equally similar, although the latter pair seems a lot closer "from the point of view of shared included sub-sets" (Morris 1979-80:456 *n* 3).

tent later on, however. In (1987:105) he is in agreement with Forte, the latter suggesting that ic and pc inclusion are together effective indicators of similarity. Forte, when discussing his compound relations,[15] states that *maximum similarity with respect to both pitch class and interval class* will be regarded as more significant than pitch class similarity alone or interval class similarity alone (1973a:50. Emphasis Forte's). To get from his discussion to ours, we just replace interval-classes and pitch-classes with the single notion of subset-class, and allow all categories of these to contribute to similarity assessments.

2.4.5 Criterion C6: Observe Non-Shared Mutually Embeddable Subset-classes

The notion giving rise to the sixth criterion is formulated by Hoover as follows: the relationships based on similarity do not in any way account for those elements of a pcset not involved in the common-tone tally or the shared subset search (1984:171).

At first glance, C6 may seem to be a bit out of place in our criteria. If we are assessing the degree of similarity between two SCs with the help of their mutually embedded subset-classes, their *unilaterally* embedded subset-classes would appear to be the ones which provide the dissimilarity. This is not entirely so, however. As we compare many pairs of SCs, there is no way we can assume that the relationship between the non-shared subset-class materials is a constant from case to case. That is, we cannot automatically equate non-shared with *completely* dissimilar.

Suppose we have two septad classes, 7-X and 7-Y. When comparing the two, our intuitive starting point is probably something like this: 7-X and 7-Y are different objects, but there is a degree of similarity between them. In order to determine the degree, we decide to compare their entire subset-class contents. The first step is to determine their subset-classes of cardinality six. Let the hexad classes included in 7-X be instances of 6-A, 6-C and 6-E, and the hexad classes included in 7-Y instances of 6-A, 6-B and 6-D.[16]

The only shared subset-class is 6-A. The unilaterally embedded classes are 6-C and 6-E in 7-X, and 6-B and 6-D in 7-Y. It would appear, then, that the hexad class contents of 7-X and 7-Y are rather different. And indeed they could be, as we

---

[15] Combinations of Forte's individual similarity relations. See section 3.5.5.

[16] The number of instances of hexad classes included in a septad class is seven. We do not need to determine specifically how many instances of each hexad class are embedded in the superset-class, but let us assume that the hexad classes are represented approximately evenly.

could deem 6-C and 6-E strongly dissimilar to 6-B and 6-D. But this is not by any means the only possibility, as we could also deem 6-C and 6-E strongly *similar* to 6-B and 6-D. Suppose, for example, that all four classes are some "near-chromatic" hexad classes with a strong sense of similarity between them. Or, under T_n-classification, we could even find out that both classes in the {6-C,6-E} pair have their inversionally related classes in the {6-B,6-D} pair.

We could ignore the meaning of the potentially high degree of similarity between these two pairs of hexad classes for an obvious reason. They are not mutually embedded. But in a sense it would be illogical to do so: when relating the two septad classes we saw that they are different objects with a degree of similarity. And now, when relating hexad classes, we would deny the importance of the very same principle only because we are dealing with subset-classes. The classes in the pair {6-B,6-C}, for example, are different objects with a degree of similarity as well, and that degree has an effect on the degree of similarity between 7-X and 7-Y. If we do not somehow take into account the unilaterally embedded classes, we are about to design a similarity measure that equates non-shared with totally dissimilar. We believe that it is a mistake to do so.

Observing mutually embedded subset-classes is only the first step in comparing two SCs. It amounts to registering the extent to which two subset-class contents consist of same elements. After this, we have to (a) assess the degree of similarity between the two groups of unilaterally embedded subset-classes, and (b) relate this degree to the first one produced by the mutually embedded classes. The final result, hopefully, corresponds more truthfully to the degree of similarity between 7-X and 7-Y than the shared subset-class count alone. There is no single definitive way to take and combine these steps, of course, but RECREL will offer one alternative.

It is evident that the sixth criterion is of relevance only for similarity measures comparing entire subset-class contents. When comparing ic contents, there is no meaningful way to do any further processing to the unilaterally embedded materials.[17] Unintentionally taking the unilaterally embedded instances into account, as measures which square component differences do, is not the same thing as fulfilling C6.

---

[17] We can, of course, envisage some methods of relating also the unilaterally embedded ic materials, for example by weighting each ic instance with some factor reflecting a degree of dissonance, etc. Approaches like these, however, are outside the scope of the present study.

2.5 STATUS OF DIFFERENT SUBSET-CLASS CARDINALITIES

If a similarity measure utilizes subset-classes of all cardinalities and thus fulfills criterion C5, two further notions immediately suggest themselves. The first is how the measure processes the subset-classes: either as one large entity, or with the different cardinalities as separate entities.

Suppose we compare two 9-pc SCs with a measure belonging to the former category.[18] There are 9 instances of 8-pc classes embedded in both nonad classes, against 126 pentad class and 126 tetrad class instances. As the measure gives an equal footing to every subset-class instance regardless of its cardinality, it seems obvious that the octad class instances are too few to affect the outcome very much, even if their contribution points strongly to similarity or dissimilarity. Such a measure has a built-in tendency to emphasize the subset-class cardinalities with the largest numbers of instances.[19] Other measures, like RECREL and its earlier version, T%REL, adopted the alternative approach. The argument goes that because we cannot demonstrate that a given subset-class cardinality is inherently more important than the others, we examine them separately and give an equal share for each partial result.

The other notion of central importance for criterion C5 is whether subset-classes of different cardinalities reflect similarity in a consistent manner.[20] Suppose, for example, that we examine some SCs X and Y by performing pair-wise comparisons to their 2CVs, 3CVs, 4CVs, etc., up to as high a subset-class cardinality as the current SCs allow. If the measure we use suggests that 2CV(X) and 2CV(Y) are highly similar, do we expect that 3CV(X) and 3CV(Y), 4CV(X) and 4CV(Y), etc., will be highly similar, too? Or, despite the dyad class similarity, could other vector comparisons indicate high dissimilarity? If so, which one of the results is to be trusted?

Let us start our analysis by observing the highly different numbers of SCs in the different cardinality-classes. The number of dyad classes, for example, is six, the number of hexad classes under T_n-classification 80. As the number of cardinality-class members equals the number of indexes in the corresponding n-class vector, the different values for n indicate vectors of radically different

[18] The category includes three measures we will examine in chapter three, TMEMB, ATMEMB and REL.

[19] This is not to suggest that we contest the validity of the measures with this feature. Some of them have obvious merits.

[20] Isaacson discusses related topics in (1992:99-106).

lengths.[21]

From the point of view of nonzero components, some vectors can be a lot more "sparse" than others. For example, the sum of components in a 6-class vector of a septad class is just 7, meaning that more than 90% of the 80 components are zeros. As the 6CVs of all septad classes are level in this manner, they appear to be highly similar: for a given component 0 in a given 6CV, the corresponding component in another 6CV is in all likelihood also 0.[22] The resemblance is evident, but concluding SC similarity from this would be a serious mistake: the contours of the vectors are similar, not the subset-class contents of the septad classes. Corresponding zero components are at best only indirect indicators of similarity, revealing what is mutually excluded. Only nonzero components participate in describing the subset-class contents of a SC, and, consequently, their status is completely different from that of the zero components. A measure comparing vectors must be designed with this in mind.[23]

It would seem that it is more improbable to find similar distributions of nonzero components in two "sparse" vectors than in two "dense" ones, the latter referring to vectors containing mostly or solely nonzero components. In other words, we might suspect that the 6CVs of septad classes are on average more dissimilar than their 2CVs. This is indeed so. Vectors in some categories are inherently more different than those in others, an observation of importance for the total measures.

Let us briefly examine this by analysing concrete results. Four comparison groups, #7/#7, #8/#8, #9/#9 and #10/#10, were selected as test materials. Each SC pair in each group was compared twice with the $\%REL_n$ measure, first with $\%REL_2$, comparing 2-class vectors, then with $\%REL_6$, comparing 6-class vectors.[24]

---

21 Under T_n-classification, to be used in the present discussion, the numbers of SCs in the cardinality-classes, and, accordingly, the numbers of indexes in the corresponding vectors, are as follows: 2-pc SCs: 6; 3-pc SCs: 19; 4-pc SCs: 43; 5-pc SCs: 66; 6-pc SCs: 80; 7-pc SCs: 66; 8-pc SCs: 43; 9-pc SCs: 19; 10-pc SCs: 6; 11-pc SCs: 1; 12-pc SCs: 1.

22 For discussion on a probabilistic approach to subset-class contents, see Lewin (1977) and (1979-80a).

23 Even if a zero component indicates "is not represented" or "does not participate in the subset-class contents," it can still be informative, at least in a short vector. For example, the number of dyad classes is so limited that each of them represents something concrete in the mind of the observer. One glance at a dyad class vector reveals both what is present and what is absent.

24 $\%REL_n$ compares proportional subset-class distributions. Given some SCs X and Y, the $\%REL_3$ value 10 would indicate that only 10% of the triad class instances in X cannot be paired with counterpart instances in Y, and vice versa. This suggests a high degree of similarity. (Reversely, 90% of the two triad class contents correspond). Value 0 would mean that the distributions are identical. See Definitions entry "$\%REL_n$" and section 3.4.2 for more details.

In the $\%REL_2$ value group #7/#7, the average is 12. As the similarity-dissimilarity continuum ranges from 0 to 100, this figure suggests that on average, the proportional dyad class distributions of septad classes are quite similar.[25] In the $\%REL_6$ value group #7/#7, by contrast, the average value is an extremely high 93. In most cases, the hexad class contents of two septad classes have very little in common. Even the lowest individual figure in this value group is as high as 43. The SCs with this value, 7-31A and 7-31B, have four inversionally symmetric hexad classes, 6-Z13, 6-Z23, 6-Z49 and 6-Z50, mutually embedded in them.

When the comparison group cardinalities increase, the averages of the $\%REL_2$ and $\%REL_6$ value groups decrease. For comparison group #8/#8, the $\%REL_2$ average is 8, the $\%REL_6$ average 72. For the comparison group #9/#9, the corresponding figures are 5 and 40, for the comparison group #10/#10, 2 and 20.

Results like these show that different n-class vectors can indeed reflect similarity between two SCs very differently.[26] This could be seen as an observation calling into question the validity of the total measures: we may intuitively deem two SCs highly similar, but comparisons between their certain nCVs may suggest high dissimilarity instead.

Let us examine this in closer detail with a concrete example. The SCs to be compared are the inversionally related 5-pc classes 5-Z18A and 5-Z18B. Their prime forms are {0,1,4,5,7] and {0,2,3,6,7}, respectively. The 4-pc classes embedded in them are as follows:

EXAMPLE 2.2: The 4-pc subset-classes of 5-Z18A and 5-Z18B.

```
5-Z18A: 4-7, 4-12B, 4-14B, 4-16A, 4-18A
5-Z18B: 4-7, 4-12A, 4-14A, 4-16B, 4-18B
```

Only one of the subset-classes, 4-7, is embedded in both pentad classes. $\%REL_4$(5-Z18A,5-Z18B) = 80. The value points strongly to dissimilarity, a suspicious result considering the strong intuitive closeness between 5-Z18A and 5-Z18B.

It is exactly here that we see in a very concrete manner how criterion C5, telling us to compare all mutually embeddable subset-classes, and criterion C6, telling us to observe also the unilaterally embedded subset-classes, are connected.

---

[25] Inversionally related classes produce always the minimum value 0. As a result, the average is slightly lower than it is under T_n/I-classification. In this value group, the maximum is 29.

[26] The results are only a few examples from extensive comparisons between different $\%REL_n$ value group averages. Detailed descriptions of these comparisons or their results are not presented here, as they would not add anything new to the already evident conclusion.

The former produces counterintuitive results without the latter. When we examine the two groups of unilaterally embedded 4-pc SCs in Ex. 2.2, we see that each class in one group has its inversional counterpart in the other. This observation instantly links the two groups much closer together. Furthermore, some other cross-related SC pairs have a high degree of similarity between them as well. For example, comparing the pairs {4-12B,4-18B} and {4-14B,4-16B} with $\%REL_2$ produces instances of the value 17, which is the second lowest one in the entire value group #4/#4 for this measure. The lowest, 0, is produced by inversionally related and Z-related classes. The two groups of unilaterally embedded tetrad classes, seeming at first to contribute to dissimilarity only, turn out to be closely related after all. The tetrad class distributions do correlate with the intuitively experienced similarity between 5-Z18A and 5-Z18B.[27]

2.6 SIMILARITY RELATIONS AND SET-CLASSIFICATIONS

There are no evident guidelines governing the relations between a similarity relation and a set-classification. A given type of set-classification is not inherently better or worse than any other, its validity depending only on how well its generality or particularity fulfils the needs of a given task.[28] In fact, even the order between a similarity relation and a set-classification is not by any means determined. We may set our minds on a given classification and start looking for a relation most suitable for it. Or we may set our minds on a certain type of similarity relation and start looking for a classification most suitable for it.

Still, the relation between a relation and a classification is of importance. Once the classification is chosen, the universe of the objects to be compared is determined. A whole category of comparisons that is perhaps of special interest under one classification may not even exist under another.[29] Prior to selecting a classification, theoretical observations cannot point objectively to a given alternative as they can be interpreted in exactly opposite ways. The starting point can be, for example, the fact that the ic contents of a pcset are preserved under inversion. For the T_n/I-minded this is the very argument supporting inclusion of in-

---

[27] At this point we will not process the pentad classes or the two groups of unilaterally embedded tetrad classes any further. The way to get a final RECREL value out of many intermediate $\%REL_n$ values will be described in Chapter 4.

[28] For a general discussion on set-classifications, see, for example, Morris (1982) and (1987:78-84).

[29] The Z-related pairs, for example, can be compared under T_n/I-classification but not under the one provided in Forte (1964). In the latter, each Z-pair comprises a single SC.

version in the group of transformations defining a set-class. For the theorist inclined to T_n-classification it merely suggests that non-equivalent objects can share important characteristics.

T_n/I-classification is by far the most widely used type, providing the framework under which practically all similarity relations are discussed. Exceptions are relatively few (Regener 1974, Solomon 1982, Rahn 1989). RECREL will be used under T_n-classification. We want inversionally related SCs to be independent objects for the simple reason that we experience them to be different. More important still, we experience that the degree of similarity between pairs of inversionally related classes is not a constant. Some inversionally related SCs may give the impression of being extremely close to each other, while others can seem more distant. RECREL was designed with these observations in mind. It, as well as some other measures to be examined, produces values supporting these observations.

■ CHAPTER 3
EVALUATING SIMILARITY RELATIONS

3.1 INTRODUCTION

In this chapter we shall evaluate a number of previously presented similarity relations, using the criteria we defined in the previous chapter. RECREL will be introduced and analysed separately in following chapters. Some preliminary discussion will precede the actual evaluations. First, in section 3.2, we examine different possibilities in categorizing the relations. In section 3.3 we discuss a number of notions relevant to the evaluations. Among these are the order in which the relations are to be examined, conventions concerning formal notations, etc.

3.2 ONE-TO-ONE AND ONE-TO-MANY CORRESPONDENCE

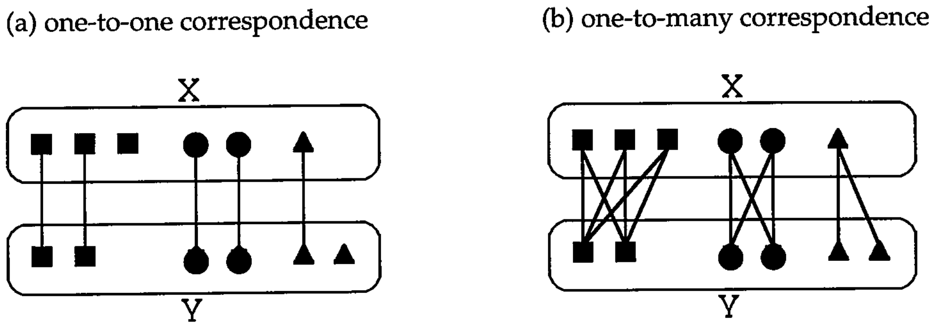
In the previous chapters, we have already identified a few categories that a vector-based similarity relation can represent. It can be a "plain" similarity relation not producing numeric values, a similarity measure comparing one subset-class cardinality at a time, or a total measure comparing entire subset-class contents. We will examine two more categories needed during the evaluations. They apply to the similarity measures only. This time, the decisive factor is not the number of compared subset-class cardinalities. Instead, we examine how the subset-class instances in one SC are paired with those in another, by *one-to-one* or *one-to-many* correspondence.[1] To refer

[1] The notion of pairing by one-to-one correspondence is from Rahn (1979-80:489).

to the two categories we will use terms such as "measures based on one-to-one correspondence," and "one-to-many correspondence measures."

The two concepts are illustrated in Diagram 3.1. Let X and Y be two SCs and the squares, circles and triangles in them instances of three subset-classes. The same SC pair is given twice. (a) illustrates one-to-one correspondence, (b) one-to-many correspondence.

DIAGRAM 3.1.



In (a) we identify a subset-class instance in X and assign an instance of the same class to it from Y. The crucial point is that after being paired, the two instances do not participate in the measurement any more. We repeat this until each instance in X having a free counterpart in Y has been paired with it. From the number of instances without counterparts we can then infer the extent of deviation from the ideal, the pairing of all instances. From the point of view of the non-paired X-square and Y-triangle, it does not matter at all whether or not the other SC contains any squares or triangles. They are surplus instances anyway. All measures taking the differences between corresponding vector components are examples of one-to-one correspondence measures.

Pairs of corresponding instances are formed also in (b). This time, however, the instance pairs are not distinct. Assigning a counterpart for a Y-square, for example, does not prevent it from being paired with the two other X-squares as well. In this approach we assume that each pair stands out independently, contributing something to similarity. The fact that the sole X-triangle is paired with the Y-triangle to the left does not mean that it ceases to exist from the point of view of the Y-triangle to the right. The latter pairing indicates "triangle similarity" just as well. Measures taking the products of corresponding vector components are examples of one-to-many correspondence measures.

It is not possible to analyse the descriptive powers of the two approaches without examining SC pairs with measures from both categories and comparing the results. This, in turn, has to wait until the measures have been properly discussed. Some general observations can be made, however. The one-to-many correspondence measures equate an increasing number of pairings with increasing similarity without paying attention to how the pairs are brought about. If in Diagram 3.1 (b) X had one triangle and Y nine, or if both X and Y had three, the result would be the same. This approach is in a sense an atomistic one, not relying on resemblances between instance distributions in X and Y, but on large numbers of individual tokens of similarity. For the one-to-one correspondence measures, where the criterion of similarity is similar amounts of similar elements, distributional resemblance is the main point.

3.3 ABOUT THE EVALUATIONS

The similarity relations will be grouped so that related constructs are adjacent. There are three categories, given in an order that in our opinion reflects increasing descriptive powers: (1) "plain" similarity relations not producing values, (2) measures processing one subset-class cardinality at a time, (3) total measures. The one-to-one / one-to-many correspondence categorization, then, is not going to be used as a basis for grouping the relations.

There will be one exception from this strength category grouping, however. Two measures belonging to the second category, $\%REL_n$ and the k measure, will be presented first. (Section 3.4). The reason for this is purely practical. We will often use the two as points of reference when discussing other measures, and it would be inconvenient to refer constantly to concepts not yet introduced.

The first category contains similarity relations by Forte, Alphonce and Solomon. (Section 3.5). The second contains measures by Teitelbaum, Morris, Lord, Isaacson, Rogers and Rahn (section 3.6), and the third the total measures by Rahn, Lewin and Castrén. (Section 3.7). Within each category, the order is chronological, again with an exception. In the second category, Rahn's ak and $MEMB_n$ will be examined last since they and the two expanded versions of the latter, TMEMB and ATMEMB, were originally presented together. TMEMB and ATMEMB will be the first total measures in the third category, and it is convenient to introduce their basic principles using the most straightforward version, $MEMB_n$.

Each similarity relation will be examined in its original context. For example, if a relation was introduced using T_n/I-classification, we will use that classification as well. If a given relation was meant to relate SCs of the same cardinality only, we will not use it to compare SCs of different cardinalities, and so on. Applying the total subset-class contents criterion C5 and the non-common subset-class criterion C6 to relations which compare only one subset-class cardinality at a time is something of a borderline case. On the one hand, some theorists have expressly stated that their measures were meant to process only limited materials, like ic contents. Applying C5 and C6 would then mean examining the measures outside their original context. On the other hand, we evaluate and compare theoretical constructs, not the opinions behind them. We want the evaluations to reflect our conviction that entire subset-class contents are better indicators of SC similarity than ic contents. Therefore, C5 and C6 will be applied to every relation.

If a measure allows comparisons between SCs of different cardinalities, the entire value group #2-#12/#2-#12 is obtained in order to calculate minima, maxima, averages and numbers of distinct values. Under T_n/I-classification, this value group contains 24,531 values. Minimum, maximum and average values and numbers of distinct values are also calculated separately for each value group #n/#m, $3 \leq n,m \leq 9$. The results are given in the 16 value group information tables.

For the sake of consistency we will adopt some representational conventions. At times, then, our expressions and notations can differ from those given by writers discussing their own relations. For instance, the original source might illustrate a measure with the help of pcsets, while we will do this with the help of SCs. Some of our conventions will follow those in Isaacson (1990). Let X and Y be two SCs, ICV(X) and ICV(Y) their interval-class vectors, respectively, and x_i and y_i the components in index i in ICV(X) and ICV(Y), respectively. The expression #ICV(X) indicates the sum of components in ICV(X) and #nCV(X) the sum of components in the n-class vector of X. X_C and Y_C are the complement classes of X and Y, respectively. X and I(X) are inversionally related classes. The function MIN returns the smaller of two numbers.

3.4 TWO REFERENCE MEASURES

3.4.1 Morris: The K Measure

Presented in Morris (1979-80:448). A similarity measure pairing interval-class instances by one-to-one correspondence.[2]

COMPARISON PROCEDURE:

The smaller components in each pair of corresponding ICV components are added together.

EQUATION:[3]

Given SCs X and Y and the function MIN,

$$k(X,Y) = \sum_{i=1}^{6} \mathrm{MIN}(x_i - y_i)$$

EXAMPLE 3.1: k(5-1,5-16).[4]

```
5-1:  {0,1,2,3,4}, [4 3 2 1 0 0]
5-16: {0,1,3,4,7}, [2 1 3 2 1 1]
                MIN 2 1 2 1 0 0
```

k(5-1,5-16) = 2+1+2+1+0+0 = 6

EVALUATION CRITERIA FULFILLED:

C1, C2, C3.3, C3.4, C4.

THE K MEASURE VALUE GROUP #2-#12/#2-#12: [5]

All values are integers. Value indicating highest degree of similarity: 55. Value indicating highest degree of dissimilarity: 0. Average: 10. Number of distinct values: 35. k(X,Y) may or may not be $k(X_C,Y_C)$.

---

[2] Morris does not give k as a similarity measure. The use as such was suggested in Rahn (1979-80). The term *measure* in connection with k was adopted for the purposes of this study.

[3] The k equation Morris gives in (1979-80) is entirely different from the one given here. Morris derives k values from those produced by the SIM measure. We define k as an independent measure. The equation is from Isaacson (1992:42).

[4] The example is from Rahn (1979-80).

[5] As the measure does not meet the value commensurability criterion C3.1, this information is of limited importance.

TABLE 3.1: The k measure value groups #n/#m, 3 ≤ n,m ≤ 9. Within a given value group, higher val - ues indicate higher degrees of similarity. Each table cell contains, clockwise from the top left: the lowest and highest values, the number of distinct values, the average.

	#3	#4	#5	#6	#7	#8	#9
#3	0 2 0.95 3						
#4	0 3 1.91 4	1 6 3.31 6					
#5	1 3 2.6 3	2 6 4.86 5	4 10 7.3 7				
#6	1 3 2.86 3	2 6 5.6 5	4 10 8.98 6	6 15 11.95 10			
#7	3 3 3.0 1	4 6 5.95 3	7 10 9.85 4	9 15 14.15 6	15 21 18.3 7		
#8	3 3 3.0 1	6 6 6.0 1	10 10 10.0 1	11 15 14.9 5	17 21 20.56 5	23 28 25.31 6	
#9	3 3 3.0 1	6 6 6.0 1	10 10 10.0 1	15 15 15.0 1	21 21 21.0 1	25 28 27.78 4	33 35 33.95 3

3.4.1.1 Analysis

The k measure counts numbers of ic instances embedded mutually in two SCs. It is a sort of opposite of the SIM measure, presented also in Morris (1979-80). The latter counts numbers of unilaterally embedded ic instances. (Section 3.6.2). Morris did not give k as a way of assessing the degree of similarity between SCs. This function was adopted by Rahn, who deems k superior to SIM (Rahn 1979-80:488-9). The two measures are closely connected, as k values can be calculated from known SIM values, and vice versa. Given SCs X and Y and the sums of their ICV components #ICV(X) and #ICV(Y), respectively, [6]

(1) k(X,Y) = (#ICV(X) + #ICV(Y) - SIM(X,Y))/2.

(2) SIM(X,Y) = (#ICV(X) + #ICV(Y) - 2k(X,Y)).

Given SCs X and Y of cardinalities n and m, respectively, the maximum value in the value group #n/#m is usually the smaller of #ICV(X) and #ICV(Y) (Rahn 1979 - 80:489). However, if n = m and there are no Z-related SCs in the cardinality-class, this maximum value can occur only if X = Y. As our value group information tables are compiled so that a SC is never compared to itself, the X = Y values do not contribute to Table 3.1. In value groups #3/#3 and #9/#9, the maximum value is #ICV(X)-1.

[6] (1) is from (Morris 1979-80:448). (2) is from (Rahn 1979-80:489).

The k measure will be used in this study to identify comparable cases when assessing the relation between a measure and the criterion C4. To make the values more informative, we will modify the notation slightly and associate each value with the maximum value in the corresponding value group. We will write k(X,Y) = n(m), where n is the number of mutually embedded ic instances between X and Y, m the maximum value. For example, k(6-1,6-32) = 11(15) indicates that the number of mutually embedded ic instances between the hexad classes is 11, and that the maximum value in the k value group #6/#6 is 15 (between Z-related classes). Likewise, the k value in example 3.1 is renotated 6(10).

Among the criteria which k fails to meet is the value commensurability criterion C3.1. As the scale of values is not the same for all value groups, a given value v can indicate many different degrees of similarity. For example, in all value groups #3/#m, $4 \leq m \leq 9$, the value indicating the highest degree of similarity is 3. Along the lines of the discussion in section 2.4.2.1, we deem these values incomparable.

Table 3.1 shows that k offers a poor degree of discrimination for a number of value groups. The measure suggests, for example, that all nonad classes are equally similar to all hexad classes (value 15). Also the value group #9/#7 consists of instances of only one value, 21. As k and SIM values were seen to be closely correlated, it is not surprising that a uniform-valued value group for k means a uniform-valued value group also for SIM. In fact, the number of distinct values is always the same in corresponding k and SIM value groups. (Value group information tables 3.1 and 3.4).

Due to the correlation between k and SIM, other important aspects of their descriptive powers also correspond and can be examined together. This will be done in connection with SIM, in section 3.6.2.2. The reason for postponing the analysis at this stage is that we will identify aspects of SIM with the help of yet another similarity measure, $\%REL_n$. The two measures are based on highly similar comparison methods. We will then examine how the results of the SIM analysis can be applied to k.

3.4.2 Castrén: $\%REL_n$

Percentage Relation. A modification of Lord's *sf* measure. Presented in an unpublished manuscript (1990). A similarity measure comparing proportionate subset-class contents, one subset-class cardinality at a time. A one-to-one correspondence measure.

COMPARISON PROCEDURE:

The sum of the absolute values of the differences between corresponding components in two n-class %-vectors is divided by two. The value is rounded to the nearest integer.

EQUATION:

Given SCs X and Y, their n-class %-vectors nC%V(X) and nC%V(Y) of length p, and components x_i and y_i in the index i of the vectors,

$$\%REL_n(X,Y) = \frac{\sum_{i=1}^{p} |x_i - y_i|}{2}$$

EXAMPLE 3.2: $\%REL_2$(6-1,6-Z4). Prime forms, 2CVs and 2C%Vs.[7]

```
6-1:  {0,1,2,3,4,5}, [5 4 3 2 1 0], [33 27 20 13  7 0]
6-Z4: {0,1,2,4,5,6}, [4 3 2 3 2 1], [27 20 13 20 13 7]
```

$\%REL_2$(6-1,6-Z4)=

$$\frac{|33-27|+|27-20|+|20-13|+|13-20|+|7-13|+|0-7|}{2} = \frac{6+7+7+7+6+7}{2} \approx 20$$

EXAMPLE 3.3: $\%REL_3$(6-1,6-Z4). 3CVs and 3C%Vs.

```
3CV(6-1) =[4 6 4 2 0 2 2 0 0 0 0 0],3C%V(6-1) =[20 30 20 10  0 10 10  0 0 0 0 0]
3CV(6-Z4)=[2 2 4 4 2 2 2 2 0 0 0 0],3C%V(6-Z4)=[10 10 20 20 10 10 10 10 0 0 0 0]
```

$\%REL_3$(6-1,6-Z4)=

$$\frac{|20-10|+|30-10|+|20-20|+|10-20|+|0-10|+|10-10|+|10-10|+|0-10|}{2}$$

$$= \frac{10+20+0+10+10+0+0+10}{2} = 30$$

EVALUATION CRITERIA FULFILLED:

C1, C2, C3.1, C3.2, C3.3, C3.4, C4.

THE $\%REL_2$ VALUE GROUP #2-#12/#2-#12:

All values are integers. Value indicating highest degree of similarity: 0. Value indicating highest degree of dissimilarity: 100. Average: 30. Number of distinct values: 85. $\%REL_2(X,Y)$ may or may not be $\%REL_2(X_C,Y_C)$.

---

[7] $\%REL_n$ will be examined under T_n/I-classification. As all n-class vectors are now involved, we will use the term 2CV instead of ICV.

TABLE 3.2: The $\%REL_2$ value groups #n/#m, $3 \leq n,m \leq 9$. Lower values indicate higher degrees of similarity. Each table cell contains, clockwise from the top left: the lowest and highest values, the number of distinct values, the average. C = 100.

	#3	#4	#5	#6	#7	#8	#9
#3	33 C 68.35 3						
#4	0 C 58.64 7	0 83 44.8 6					
#5	13 90 57.12 15	10 80 38.85 20	0 60 27.0 7				
#6	13 87 57.02 12	10 80 36.61 20	0 60 24.63 18	0 60 20.33 10			
#7	29 86 56.81 13	7 76 34.5 26	4 57 21.72 32	4 57 17.65 38	0 29 13.01 7		
#8	36 86 56.8 15	6 75 33.45 42	6 57 21.05 38	4 57 16.63 35	0 29 11.87 21	0 18 9.66 6	
#9	42 83 56.75 14	6 75 32.23 23	7 56 20.0 32	5 56 15.91 29	4 27 10.58 19	4 18 8.59 13	3 8 5.94 3

3.4.2.1 Analysis

Although $\%REL_n$ was originally designed to be an independent similarity measure, its main function is to be an integral part of RECREL. It is an internal measure of similarity in the latter, evaluated hundreds or even thousands of times during a single RECREL comparison between two SCs of large cardinalities.

$\%REL_n$, unlike Lord's Similarity Function, from which it is derived, is not restricted to comparisons between SCs of the same cardinality. Nor is it restricted to comparisons between dyad classes only. Any two n-class vectors belonging to some SCs X and Y can be transformed into n-class %-vectors. nC%V(X) and nC%V(Y), then, give the proportionate n-class contents of X and Y, respectively, and the corresponding $\%REL_n$ value give the degree of dissimilarity between those contents.

E.g. the lowest value, 0, is produced by $\%REL_2$ comparisons between Z-related SCs, between inversionally related SCs under T_n-classification, and between SCs having different 2CVs but identical 2C%Vs. E.g. $\%REL_2$(6-35,5-33) = 0, meaning that the proportionate dyad class contents are identical. The 2CVs and 2C%Vs are given in Ex. 3.4.

EXAMPLE 3.4: 2-class vectors and 2-class %-vectors of 6-35 and 5-33.

```
2CV(6-35) = [0 6 0 6 0 3], 2C%V(6-35) = [0 40 0 40 0 20]
2CV(5-33) = [0 4 0 4 0 2], 2C%V(5-33) = [0 40 0 40 0 20]
```

Among the few other set-class pairs whose %REL_2 value is 0, one finds {3-10,4-28}, {3-8,4-25}, {3-5,4-9}, {5-21,6-20} and {7-31,8-28}. One of the SCs in these pairs produces exactly the same %REL_2 values with the rest of the classes as its counterpart does.

The highest %REL_n value, 100, indicates that the two SCs being compared do not share any subset-classes of cardinality n. For example %REL_2(4-9,3-6) = 100. The vectors are in Ex. 3.5.

EXAMPLE 3.5: 2-class vectors and 2-class %-vectors of 4-9 and 3-6.

```
2CV(4-9) = [2 0 0 0 2 2], 2C%V(4-9) = [33 0 0  0 33 33]
2CV(3-6) = [0 2 0 1 0 0], 2C%V(3-6) = [0 67 0 33  0  0]
```

As %REL_n compares only one cardinality-class at a time, it does not meet the total subset-class contents criterion C5. Nor does it meet the non-common subset-class criterion C6. When n exceeds 2, %REL_n can to a certain extent discriminate between inversionally related SCs under T_n-classification. Given an inversionally non-symmetric SC X, in most cases %REL_n(X,I(X)) > 0 when n > 2. The outcome is not certain, however. For example, %REL_3(6-14A,6-14B) = 0. The 3-class %-vectors of the two SCs are identical. Ex. 3.6.

EXAMPLE 3.6: 3C%V(6-14A) = 3C%V(6-14B).

```
[5 5 5 10 10 10 10 0 0 5 5 5 0 0 5 0 10 10 5]
```

Perhaps the main advantage in comparing proportional subset-class distributions is that we get an intuitively satisfactory gradation of similarity in cases that are equally similar from the point of view of absolute subset-class distributions, as described by the n-class vectors. Again, we have to postpone a more detailed analysis until a suitable point of reference, the SIM measure, has been presented. %REL_n and SIM are compared in section 3.6.2.2.

Oddly enough, the proportional distributions also cause a problem for %REL_2. A 2C%V does not reveal the cardinality of the SC it belongs to. As a result, cardinalities as factors contributing to similarity assessments cease to matter. If the 2-class %-vectors of two SCs are highly similar, %REL_2 deems the SCs highly similar, even if one is a tetrad class and the other a nonad class. The value group information table 3.2 shows that the lowest value in the value group #4/#9 is 6, i.e., an extremely low value. Furthermore, the average value in the value group #9/#5 is 20, lower than the value group #5/#5 average. The reason is that all nonad classes and many pentad classes have reasonably level 2-class %-vectors, resulting in a relatively low average. Some pentad classes, by contrast, have peaked 2%CVs, and pairwise comparisons between these increase the #5/#5 average.

Generally, some of the comparison groups with highly different cardinalities seem to produce counterintuitively low values. We could, of course, defend $\%REL_2$ by saying that it measures truthfully what it was set to measure. A more fruitful approach, however, is to seek for improvements by applying the total subset-class contents criterion C5. By comparing entire subset-class contents the cardinality difference will be reflected in the final result, even if the proportional dyad class distributions would be highly similar. T%REL, the resulting expansion of $\%REL_n$, is to be examined in section 3.7.4. RECREL, in turn, is the result of applying the non-common subset-class criterion C6 to T%REL.

3.5 SIMILARITY RELATIONS NOT PRODUCING NUMERIC VALUES

3.5.1 Forte: The R_1 Relation

Presented in Forte (1973a:46-60). A similarity relation comparing interval-class vectors. No numeric values produced. The outcome is an indication whether or not the relation holds. Applies to SCs of the same cardinality only.

COMPARISON PROCEDURE:
Two SCs are in the relation R_1 if four of the six corresponding components in their ICVs are the same and the remaining two correspond crosswisely.

EXAMPLE 3.7: R_1(4-2,4-3). Prime forms and ICVs.[8]

```
4-2: {0,1,2,4}, [2 2 1 1 0 0]
4-3: {0,1,3,4}, [2 1 2 1 0 0]
```

The corresponding components in the two vectors are of the same size in four indexes, 1, 4, 5 and 6. In indexes 2 and 3 the corresponding components are different. The two latter pairs consist of the same components, 1 and 2. 4-2 and 4-3 are in the R_1 relation.

---

[8] Examples 3.7 and 3.8 are from Forte (1973a:48).

EVALUATION CRITERIA FULFILLED:
None.

3.5.1.1 Analysis

In (1964) and (1973a:46-60) Forte discusses similarity relations based on comparisons between ICVs. All are intended for SCs of the same cardinality only. The limitation is no doubt intentional. It becomes evident from Forte's discussion that he sees the inclusion relation as an important element in assessing similarity between SCs (1973a:24-46, 93-108). As inclusion considerations exclude SCs of the same cardinality, the need for specialised relations emerges.

R_1 does not meet any of our criteria. Due to their similarity, R_1 and its closely related variant R_2 are examined together in section 3.5.2.1. Further aspects of Forte's relations are examined in section 3.5.4.

3.5.2 Forte: The R_2 Relation

Presented in Forte (1973a:46-60). A similarity relation comparing interval-class vectors. No numeric values produced. The outcome is an indication whether or not the relation holds. Applies to SCs of the same cardinality only.

COMPARISON PROCEDURE:
Two SCs are in the relation R_2 if four of the six corresponding components in their ICVs are the same, but the remaining two do not correspond crosswisely.

EXAMPLE 3.8: R_2(5-10,5-Z12). Prime forms and ICVs.

```
5-10:  {0,1,3,4,6}, [2 2 3 1 1 1]
5-Z12: {0,1,3,5,6}, [2 2 2 1 2 1]
```

The corresponding components in the two vectors are of the same size in four indexes, 1, 2, 4 and 6. Components in indexes 3 and 5 form pairs {3,2} and {1,2}, contents of which are not identical. The SCs are in the R_2 relation.

EVALUATION CRITERIA FULFILLED:
None.

3.5.2.1 Analysis

According to Forte, R_1 and R_2 indicate maximum similarity with respect to interval class (1973a:49), the former representing a closer relationship than the latter (Ibid., 48). Since the common criterion is equality between four pairs of corresponding components, comparisons are launched by observing similarities between ic distributions. This notion, however, is not applied to the remaining two pairs of corresponding components. They can be nearly similar or highly different without affecting the outcome. Intuitively, then, they represent clearly varying degrees of similarity, suggesting a hidden gradation of similarity for pairs enjoying R_1 or R_2. See Ex. 3.9.

EXAMPLE 3.9: R_1-related SC pair {6-1,6-32} and R_2-related SC pair {6-Z24,6-Z26}.

```
ICV(6-1)  = [5 4 3 2 1 0]     ICV(6-Z24) = [2 3 3 3 3 1]
ICV(6-32) = [1 4 3 2 5 0]     ICV(6-Z26) = [2 3 2 3 4 1]
```

The pair to the left is R_1-related, the pair to the right R_2-related. In the ICVs of the pair {6-1,6-32}, both pairs of non-identical corresponding components produce the difference four. In the ICVs of the pair {6-Z24,6-Z26}, the difference is in both cases one. One third of the indexes in the former pair contain highly different components, whereas in the latter, the ic contents represent the smallest possible deviation from the identical.[9] R_1 and R_2, however, deem the former pair more similar. Dissimilarity between corresponding components is ignored, since greater similarity is available between *non*-corresponding components. The sole ic5 instance in 6-1 is paired with the ic1 instance in 6-32, the five ic1 instances in 6-1 with the five ic5 instances in 6-32. Correspondence like this is completely artificial. From the point of view of the size difference between components in index 1, it is irrelevant what sort of components reside in index 5. The extent to which one class of objects is present does not balance out, rectify or contradict the extent to which another class of objects is present.

[9] $\%REL_2$(6-1,6-32) = 27; $\%REL_2$(6-Z24,6-Z26) = 7; k(6-1,6-32) = 11(15); k(6-Z24,6-Z26) = 14(15).

3.5.3 Forte: The R_0 Relation

Presented in Forte (1973a:46-60). A similarity relation comparing interval-class vectors and indicating *minimum* similarity. No numeric values produced. The outcome is an indication whether or not the relation holds. Applies to SCs of the same cardinality only.

COMPARISON PROCEDURE:
Two SCs are in the relation R_0 if all corresponding components in their ICVs are different.

EXAMPLE 3.10: R_0(4-Z15,4-28). Prime forms and ICVs.

```
4-Z15: {0,1,4,6}, [1 1 1 1 1 1]
4-28:  {0,3,6,9}, [0 0 4 0 0 2]
```

All corresponding components in the two ICVs are different. The SCs are in the R_0 relation.

EVALUATION CRITERIA FULFILLED:
None.

3.5.3.1 Analysis

R_0 has the same feature as R_1 and R_2, a gradation of similarity for SC pairs enjoying it, ignored but still evident. Ex. 3.11.

EXAMPLE 3.11: R_0-related SC pairs {6-Z3,6-Z17} and {6-Z6,6-35}.

```
ICV(6-Z3)  = [4 3 3 2 2 1]    ICV(6-Z6) = [4 2 1 2 4 2]
ICV(6-Z17) = [3 2 2 3 3 2]    ICV(6-35) = [0 6 0 6 0 3]
```

Both ICVs belonging to the pair {6-Z3,6-Z17} are reasonably level, the difference between corresponding components being 1 in every index. Their k measure value is 12(15). The vectors of the pair {6-Z6,6-35}, by contrast, have considerably larger differences between corresponding components. k(6-Z6,6-35) = 6(15), which suggests a clearly lower degree of similarity than for the former pair. %REL_2(6-Z3,6-Z17) = 20; %REL_2(6-Z6,6-35) = 60.

3.5.4 R_1, R_2 and R_0: Conclusions

On the basis of the discussion above, we deem the R_1, R_2 and R_0 relations to be weak methods for demonstrating SC similarity. They have been criticised by a number of theorists (Regener 1974, Lord 1981, Chapman 1981, Isaacson 1990). Besides the coarseness of the relations and their inability to relate SCs of different cardinalities, they do not form a complete system. Many SC pairs are in none of the relations. Also, Z-related SCs are not categorised in terms of the relations (Isaacson 1990:2-4).

3.5.5 Forte: The Compound Relations

In his introductory remarks for the similarity relations in (1973a:46), Forte discusses the usefulness of relations that can determine the *degree* of similarity between two SCs. As his own relations do not offer gradation individually, he may have thought of them as constituting a "superrelation" together. Different relations would identify different levels of SC similarity, the combined result being a rough "scale" from maximally similar to minimally similar.

He does not develop this approach systematically, but offers different viewpoints to SC similarity by using his relations in a number of different combinations. These involve the R_p relation and the R_1, R_2 and R_0 relations.[10] He states that in general, maximum similarity with respect to both pitch class and interval class will be regarded as more significant than pitch class similarity alone or interval class similarity alone (1973a:50). Some of these *Compound relations*, combining R_p with R_1 and/or R_2, no doubt demonstrate similarity between SCs better than any of the relations alone. To an extent they approach the total subset-class contents criterion C5 by combining aspects of ic and pc inclusion (section 2.4.4), but they do not fare any better from the point of view of our criteria. Other compound relations are counterintuitive as they unite a relation indicating similarity (R_p, R_1 or R_2) with one indicating dissimilarity (R_0 or the *absence* of R_p). Forte does not analyse the descriptive powers of the latter compound relations at all, only calls them "extraordinary" and "interesting." He even refers to a musical example where a compound relation suggests that melody and accompaniment are "at once similar and dissimilar." (Ibid., 50-1).

[10] Forte (1973a:50-60, 182-95). For a description of the R_p relation, see (Ibid., 47-8), or section 1.1, *n* 10.

3.5.6 Alphonce: The Difference Value Relation

Presented in Alphonce (1974:149-53).[11] A similarity relation comparing corresponding n-class vectors.

COMPARISON PROCEDURE:
Absolute values of the differences between corresponding vector components constitute an *absolute-value difference vector*, or ADV.[12] The ADV components are arranged in ascending order. Nonzero numbers are interpreted as a single integer, which, in turn, is interpreted as the value of the comparison. The number of digits in the value is the same as the number of non-equal corresponding components in the compared vectors. Fewer digits indicate closer similarity.

EXAMPLE 3.12: The Difference Value between SCs 4-1 and 4-2.[13]

```
2CV(4-1) = [3 2 1 0 0 0]
2CV(4-2) = [2 2 1 1 0 0]
       ADV [1 0 0 1 0 0] => Asc.order:[0 0 0 0 1 1] => Value: 11
```

EVALUATION CRITERIA FULFILLED:
None (?).

3.5.6.1 Analysis

When introducing the Difference Value relation in (1974:149), Alphonce warns his readers that his discussion is highly tentative. He then presents the relation in such poor detail that it is difficult to assess how it relates to our evaluation criteria. It is not indicated, for example, whether SCs of different cardinalities can be compared. Alphonce suggests that any n-class vectors or even entire subset-class contents could be compared, not merely 2CVs. He offers no concrete analysis or results, however, only an observation that with other than 2-class vectors the difference values be-

11 Alphonce does not give a name to his relation, but calls its results "difference values". We will adopt this name for the whole comparison method.

12 Alphonce uses the term Difference Vector, but in this study the term is reserved to another kind of vector. The term ADV is from Isaacson (1992:111).

13 The example is from Alphonce (1974:151).

come clumsy (Ibid., 152-53).[14]

We saw above that Forte's R_1 and R_2 relations emphasize identical corresponding components and ignore the contribution of the non-identical ones. R_0, in turn, equates non-identical with minimally similar, although different cases may represent varying degrees of similarity. Alphonce's relation has defects resembling these.[15] Ex. 3.13 gives two SC pairs already familiar from examples above.

EXAMPLE 3.13: 2CVs and ADVs of SC pairs {6-1,6-32} and {6-Z3,6-Z17}.

```
2CV(6-1)  = [5 4 3 2 1 0]      2CV(6-Z3)  = [4 3 3 2 2 1]
2CV(6-32)= [1 4 3 2 5 0]       2CV(6-Z17)= [3 2 2 3 3 2]
       ADV [4 0 0 0 4 0] => 44         ADV [1 1 1 1 1 1] => 111111
```

In the 2CVs of the {6-1,6-32} pair, dissimilarities occur in only two indexes. In the other vector pair, all corresponding components differ. As a result, the first value, 44, belongs to the most similar, two-digit value category. The latter value, 111111, with its six digits, belongs to the most dissimilar one. Both %REL_2 and the k measure, however, suggest closer similarity to the latter pair. %REL_2(6-1,6-32) = 27; %REL_2(6-Z3,6-Z17) = 20; k(6-1,6-32) = 11(15); k(6-Z3,6-Z17) = 12(15). The implication that few, but dramatic, component size differences automatically indicate closer similarity than several small ones is unsubstantiated.

The relation also inadvertently takes into account something else than subset-class content similarity, i.e., it favors peaked vectors. See Ex. 3.14.

EXAMPLE 3.14: 2CVs and ADVs of four triad class pairs.

```
2CV(3-10) = [0 0 2 0 0 1]    2CV(3-7)  = [0 1 1 0 1 0]
2CV(3-12) = [0 0 0 3 0 0]    2CV(3-12) = [0 0 0 3 0 0]
        ADV [0 0 2 3 0 1]            ADV [0 1 1 3 1 0]

2CV(3-9)  = [0 1 0 0 2 0]    2CV(3-5) = [1 0 0 0 1 1]
2CV(3-10) = [0 0 2 0 0 1]    2CV(3-6) = [0 2 0 1 0 0]
        ADV [0 1 2 0 2 1]           ADV [1 2 0 1 1 1]
```

The four SC pairs are comparable in the sense that each SC is of the same cardinality and the dyad class contents of the SCs in each pair are disjoint. (The k value is 0(2)

---

[14] Alphonce also outlines an alternative for his relation, which turns out to be exactly the SIM relation presented in Morris (1979-80). He rejects it, however, on the grounds that it loses a viewpoint he deems important, the number of equal corresponding components in the compared vectors.

[15] Forte's R relations relate to Alphonce's so that R_0 corresponds to six-digit difference values, R_1 and R_2 to two-digit values. Difference values do not distinguish between R_1 and R_2 (Alphonce 1974:152).

for every pair). Yet the Difference Value relation produces a different value for every pair. The values are 123, 1113, 1122 and 11112, in order of increasing dissimilarity. Moreover, some triad class pairs with k value 1(2) have a difference value indicating a higher degree of dissimilarity than the value 123. The value between the SCs in the pair {3-1,3-5}, for example, is 1111, although the SCs share an instance of the dyad class 2-1.

The Difference Value relation is a peculiar hybrid between a measure and a non-numeric similarity relation. It does produce numeric values, but these constitute a scale where consecutive values are separated both by reasonably small intervals and by huge leaps, resulting in similarity assessments that are quite simply absurd. The crucial point is that Alphonce does not say how the values are to be interpreted, as genuine values on a genuine scale or as some other types of results identifying some approximate categories of similarity. We assume the latter alternative is the correct one, and deem the values meaningless. The former alternative would suggest that the {6-Z3,6-Z17} pair is 2,525 times more dissimilar than the {6-1,6-32} pair. The entire value group #6/#6, if taken literally, would produce even more controversial results.

3.5.7 Solomon: The R Relation

Presented in Solomon (1982). A similarity relation comparing both interval-class vectors and pc contents. No numeric values produced. The outcome is an indication whether or not the relation holds. Applies to SCs of the same cardinality only.

COMPARISON PROCEDURE:
SCs X and Y of cardinality n are in the R relation if (1) for a pcset in X there is a pcset in Y so that the two have n-1 pcs in common,[16] (2) the numeric values of the two pcs not in common are consecutive mod 12, (3) the value for k(X,Y) is equal to or greater than the value of the formula (#ICV(X) * #X)/8.[17]

---

16 The first condition is the same as Forte's R_p relation.

17 Solomon, like a number of other theorists, is in the unfortunate habit of using the term set for both pc collections and set-classes. He defines the relation with the help of individual pc collections, the prime forms, without indicating how the relation holds abstractly between SCs. It is assumed here that the R relation, like the R_p relation in Forte (1973a:50), can be "weakly represented" between pcsets with few or no common elements.

EXAMPLE 3.15: R(6-Z26,6-34). Prime forms and ICVs.[18]

```
6-Z26: {0,1,3,5,7,8}, [2 3 2 3 4 1]
6-34:  {0,1,3,5,7,9}, [1 4 2 4 2 2]
```

Condition (1) is satisfied as there are five pcs in common between the prime forms. Condition (2) is satisfied as the non-shared pcs 8 and 9 are consecutive. Condition (3) is satisfied, as the value for k(6-Z26,6-34), 12(15), is larger than the value for (#ICV(6-Z26) * #6-Z26)/8 = (15 * 6)/8 = 11.25. The SCs are in the R relation.

EVALUATION CRITERIA FULFILLED:
None.

3.5.7.1 Analysis

Solomon never analyses the suggested importance of the smallest possible movement between non-common pcs (condition 2), or the purpose of the formula (#ICV(X)*#X)/8 (condition 3). Due to the peculiar constant divisor in the formula, the R relation treats different cardinalities quite differently. Given the comparison group #4/#4, for example, the value of the formula is 3 and the highest k value for a tetrad class pair 6. Obviously, the relation can hold if the other conditions are met. For pairs in the comparison group #9/#9, in turn, the highest k value is 35 and the value of the formula 40.5. As the latter is higher than the former, the relation cannot hold between nonad classes. In fact, the R relation does not hold between *any* SC pairs of cardinality 8 or larger (Solomon 1982:75). So drastic a feature in a similarity relation would demand most detailed analysis and argumentation, none of which can be found in Solomon (1982). The validity of the relation is highly questionable.

3.6 SIMILARITY MEASURES COMPARING ONE SUBSET-CLASS CARDINALITY AT A TIME

3.6.1 Teitelbaum: The Similarity Index (s.i.)

Presented in Teitelbaum (1965). A similarity measure pairing interval-class instances by one-to-one correspondence. For SCs of the same cardinality only. Z-related SCs are not compared.

[18] The example is from Solomon (1982).

COMPARISON PROCEDURE:
Differences between corresponding ICV components are squared and added together. The final result is the square root of the sum.

EQUATION:

$$\text{s.i.}(X,Y) = \sqrt{\sum_{i=1}^{6}\left(x_i - y_i\right)^2}$$

EXAMPLE 3.16: s.i.(4-1,4-2). Prime forms and ICVs.[19]

```
4-1: {0,1,2,3}, [3 2 1 0 0 0]
4-2: {0,1,2,4}, [2 2 1 1 0 0]
```

s.i.(4-1,4-2) =

$$\sqrt{(3-2)^2 + (2-2)^2 + (1-1)^2 + (0-1)^2 + (0-0)^2 + (0-0)^2}$$

$$= \sqrt{1 + 0 + 0 + 1 + 0 + 0}$$

$$= \sqrt{2} = 1.41$$

EVALUATION CRITERIA FULFILLED:[20]
C3.3, C3.4.

THE SET OF VALUES:[21]
Value indicating highest degree of similarity: $\sqrt{2} \approx 1.41$. Value indicating highest degree of dissimilarity: $\sqrt{72} \approx 8.49$. Average: 2.85. Number of distinct values: 31. s.i.(X,Y) = s.i.(X_C,Y_C).

19 After Teitelbaum (1965).

20 The value managability criterion C3.3 is met if the values are given in the square root notation Teitelbaum uses, and not as irrational numbers. When examining the different value groups, non-integer values were rounded to two decimal places.

21 The set of values comprises values in the nine value groups #n/#n, $2 \leq n \leq 10$. Values produced by Z-related pairs are excluded. As the measure does not meet the value commensurability criterion C3.1, this information is of limited importance.

TABLE 3.3: The s.i value groups #n/#n, $3 \leq n \leq 9$. Within a given value group, lower values indicate higher degrees of similarity. Each table cell contains: the lowest and highest values, the average, the number of distinct values.

#3/#3	1.41	3.74	2.29	7
#4/#4	1.41	5.48	2.74	14
#5/#5	1.41	6.0	2.79	18
#6/#6	1.41	8.49	3.1	30
#7/#7	1.41	6.0	2.79	18
#8/#8	1.41	5.48	2.74	14
#9/#9	1.41	3.74	2.29	7

3.6.1.1 Analysis

In Table 3.3, the figures belonging to the value group #n/#n duplicate those belonging to the value group #(12-n)/#(12-n). Complementary comparison groups, like two individual SCs and their complements, produce identical results. All value groups share the same minimum value, whereas the maxima vary. The highest individual value, 8.49, is produced by the pair {6-20,6-35}. s.i. does not meet criterion C3.1, meaning that a given value *v* can indicate different degrees of similarity in different value groups.

The measure also fails to meet C4, the criterion calling for a uniform value for all comparable cases. The reason is the squaring of the ICV component differences. (See section 2.4.3). s.i. is one of the measures treating level and peaked vectors differently. Let us examine this with a few comparisons. (Ex. 3.17).

EXAMPLE 3.17: Interval-class vectors of five triad classes.

```
ICV(3-1)  = [2 1 0 0 0 0]
ICV(3-10) = [0 0 2 0 0 1]
ICV(3-11) = [0 0 1 1 1 0]
ICV(3-12) = [0 0 0 3 0 0]
ICV(3-9)  = [0 1 0 0 2 0]
```

3-1 does not have any ic instances in common with SCs 3-10, 3-11 or 3-12. It would be natural to expect, then, that the uniform k value 0(2) would mean also a uniform s.i. value between 3-1 and each of the three other SCs. In fact, all three values are different. s.i.(3-1,3-11) = 2.83; s.i.(3-1,3-10) = 3.16; s.i.(3-1,3-12) = 3.74. The lowest of these values is with 3-11, as its ICV has as even ic instance distribution as possible. ICV(3-10) is slightly more peaked, and ICV(3-12) contains the largest component. s.i. punishes the SCs with increasingly peaked vectors with increasing values.

Furthermore, SCs 3-1 and 3-9 have one instance of ic2 in common. Despite the fact that k(3-1,3-9) = 1(2), s.i.(3-1,3-9) = 2.83. This is exactly the same value 3-1 had with 3-11, a SC with disjoint ic contents. As the component 2 in the fifth index of ICV(3-9) is squared, the "shared-instance advantage" which 3-9 has over 3-11 loses its meaning. In extreme cases, s.i. may even suggest closer similarity to a pair with a *lower* k value. Ex. 3.18.

EXAMPLE 3.18: Interval-class vectors of four hexad classes.

```
ICV(6-7)  = [4 2 0 2 4 3]
ICV(6-34) = [1 4 2 4 2 2]

ICV(6-1)  = [5 4 3 2 1 0]
ICV(6-32) = [1 4 3 2 5 0]
```

s.i.(6-7,6-34) = 5.1 and s.i.(6-1,6-32) = 5.66. The SCs in the former pair are deemed more similar than those in the latter. Yet k(6-7,6-34) = 9(15) and k(6-1,6-32) = 11(15). The principle behind s.i. is the opposite of the one behind Forte's R_1 and R_2 (section 3.5.2.1). Forte's relations ignore large individual component size differences as long as four corresponding components are equal. s.i. ignores four equal corresponding components if large component size differences are to be found. The idea implied is that a few sharp differences override an otherwise strong distributional similarity, to the extent that the very starting point, shared-ic instance count, is relegated to secondary importance. In other words, more mutually embedded ic instances means *less* SC similarity if the unilaterally embedded ic instances are not evenly distributed, but constitute peaks in the vector. (Section 2.4.3). This is a very precise claim, and in our view a mistaken one. It is not even identified, let alone analysed, in Teitelbaum (1965).

3.6.2 Morris: SIM

Presented in Morris (1979-80). A similarity measure pairing interval-class instances by one-to-one correspondence.

COMPARISON PROCEDURE:

Absolute values of the differences between corresponding ICV components are added together.

EQUATION:

$$\text{SIM}(X,Y) = \sum_{i=1}^{6} \left| x_i - y_i \right|$$

EXAMPLE 3.19: SIM(8-9,8-1). Prime forms and ICVs.[22]

```
8-9: {0,1,2,3,6,7,8,9}, [6 4 4 4 6 4]
8-1: {0,1,2,3,4,5,6,7}, [7 6 5 4 4 2]
```

$$\text{SIM}(8\text{-}9,8\text{-}1) = |6-7|+|4-6|+|4-5|+|4-4|+|6-4|+|4-2|$$

$$= 1+2+1+0+2+2 = 8$$

EVALUATION CRITERIA FULFILLED:

C1, C2, C3.3, C3.4, C4.

THE SIM VALUE GROUP #2-#12/#2-#12:[23]

All values are integers. Value indicating highest degree of similarity: 0. Value indicating highest degree of dissimilarity: 65. Average: 13. Number of distinct values: 44. If #X = #Y, SIM(X,Y) = SIM(X_C ,Y_C). If #X ≠ #Y, SIM(X,Y) may or may not be SIM(X_C,Y_C). SIM(X,X_C) = 11g/2, where g is the difference between #X and #X $_C$, #X being larger than #X_C (Morris 1979-80:451).

TABLE 3.4: The SIM value groups #n/#m, 3 ≤ n,m ≤ 9. Within a given value group, lower values indicate higher degrees of similarity. Each table cell contains, clockwise from the top left: the lowest and highest values, the number of distinct values, the average.

	#3	#4	#5	#6	#7	#8	#9
#3	2 6 4.09 3						
#4	3 9 5.19 4	0 10 5.37 6					
#5	7 11 7.8 3	4 12 6.27 5	0 12 5.4 7				
#6	12 16 12.28 3	9 17 9.8 5	5 17 7.03 6	0 18 6.1 10			
#7	18 18 18.0 1	15 19 15.09 3	11 17 11.3 4	6 18 7.69 6	0 12 5.4 7		
#8	25 25 25.0 1	22 22 22.0 1	18 18 18.0 1	13 21 13.21 5	7 15 7.89 5	0 10 5.37 6	
#9	33 33 33.0 1	30 30 30.0 1	26 26 26.0 1	21 21 21.0 1	15 15 15.0 1	8 14 8.44 4	2 6 4.09 3

22 After Morris (1979-80:447).

23 As the measure does not meet the value commensurability criterion C3.1, this information is of limited importance.

3.6.2.1 Analysis

The lowest value, 0, is produced only by comparisons between Z-related classes. (Table 3.4, value groups #n/#n, $4 \leq n \leq 8$). The scale of values is not the same in all value groups, but within a given comparison group, all comparable SC pairs have a uniform SIM value. In the comparison group #3/#4, for example, all pairs with k value 0(3) have the SIM value 9. Pairs with k value 1(3) have the SIM value 7, etc.[24]

The highest SIM values are obtained from comparisons between SCs of greatly differing cardinalities. When the difference between the comparison group cardinalities becomes sufficiently great, SIM produces value groups with multiple instances of only one value (Morris 1979-80:447-8). The uniform value is equal to #ICV(X) - #ICV(Y), #X being larger than #Y. Thus, for example, all values in the value group #3/#9 are instances of the value 33, as the sum of components in the ICV of a nonad class is 36, and that in the ICV of a triad class 3. When the value group contains instances of more than one value, the same formula #ICV(X) - #ICV(Y) gives the lowest value in the group. For example, in the value group #4/#5, the lowest value is 4. For every pentad class X, #ICV(X) = 10, for every tetrad class Y, #ICV(Y) = 6 (Ibid., 447).

3.6.2.2 SIM, $\%REL_n$ and the k Measure

Let us examine some results which SIM returns from comparisons made between SCs of two different cardinalities. We will compare SC 3-5 to four tetrad classes, 4-9, 4-6, 4-5 and 4-Z15. To get a point of reference, we compare the same pairs also with $\%REL_2$. Ex. 3.20.

EXAMPLE 3.20: The ICVs and 2C%Vs of five SCs.

```
ICV(3-5)   = [1 0 0 0 1 1], 2C%V(3-5)   = [33  0  0  0 33 33]
ICV(4-9)   = [2 0 0 0 2 2], 2C%V(4-9)   = [33  0  0  0 33 33]
ICV(4-6)   = [2 1 0 0 2 1], 2C%V(4-6)   = [33 17  0  0 33 17]
ICV(4-5)   = [2 1 0 1 1 1], 2C%V(4-5)   = [33 17  0 17 17 17]
ICV(4-Z15) = [1 1 1 1 1 1], 2C%V(4-Z15) = [17 17 17 17 17 17]
```

The SIM value between 3-5 and each of the tetrad classes is the same, 3. By contrast, the $\%REL_2$ value between 3-5 and each of the tetrad classes is different.

24 The correlation between SIM and k was examined in section 3.4.1.1.

$\%REL_2$(3-5,4-9) = 0; $\%REL_2$(3-5,4-6) = 17; $\%REL_2$(3-5,4-5) = 33; $\%REL_2$(3-5,4-Z15) = 50.

For SIM, the decisive factor in gaining the uniform value is that the nonzero components in ICV(3-5) are matched with at least equal-sized components in the ICVs of the tetrad classes. The smaller contents are in a sense "swallowed" by each of the larger ones, and the distribution of the remaining ic instances in the tetrad class ICVs has no effect on the outcome.

We might note, however, that in ICV(3-5) and ICV(4-9) the nonzero components are located exactly in the same indexes and, furthermore, within each vector, the nonzero components are of the same size. Also ICV(4-6) bears obvious resemblances to ICV(3-5), as only one of its nonzero components, in index 2, has a zero counterpart in the latter. ICV(4-5) already contains instances of two ics, 2 and 4, that are not represented in ICV(3-5). Differences between ICV(4-Z15) and ICV(3-5) are even more evident, as ics 2, 3 and 4 contribute to the former but not to the latter. With respect to 3-5, the tetrad classes seem to be ordered according to increasing dissimilarity.

Ex. 3.20 illustrates what in our opinion is a serious defect in SIM. When two different-sized ic contents are compared, an important aspect is entirely ignored, i.e., similarity between what we will call the *profiles* of the contents: are they similar not only with respect to mutually embedded, but also with respect to mutually excluded ics; do certain ics dominate both contents with similar relative strengths, etc.

We believe that comparing proportional ic instance distributions instead of absolute ones is a better way to demonstrate the sense of varying similarity between 3-5 and the 4-pc classes. To put it simply, the four $\%REL_2$ values vary because the nonzero components in 2C%V(3-5) are *large*. Every non-matching feature between a tetrad class 2C%V and 2C%V(3-5) is reflected in an increasing $\%REL_2$ value. From this point of view the position of 4-9 is unique, as its 2C%V is identical to that of 3-5. Also the relative closeness of 4-6 is evident, as only one sixth of its proportional ic contents do not match those of 3-5. For 4-5 the corresponding non-matching share is one third, for 4-Z15 it is already a half.

SIM's resolution is already somewhat coarse at the point where the difference between the cardinalities is only one. When the difference grows larger, the coarseness increases accordingly. In the value group #4/#7, for instance, the lowest value is 15. Among other pairs this value belongs to {4-Z15,7-28}. The highest value is only 4 points higher, 19. This value can be obtained, for example, from the comparison between 4-28 and 7-15. The ICVs of these two SC pairs are given in Ex. 3.21. $\%REL_2$(4-Z15,7-28) = 7; $\%REL_2$(4-28,7-15) = 76.

EXAMPLE 3.21: Interval-class vectors of four SCs.

```
ICV(4-Z15)  = [1 1 1 1 1 1]
ICV(7-28)   = [3 4 4 4 3 3]
ICV(4-28)   = [0 0 4 0 0 2]
ICV(7-15)   = [4 4 2 4 4 3]
```

SIM's coarseness reaches its extreme in the uniform-valued value groups. (Table 3.4). Given such a value group, as well as the comparison group #n/#m producing it (n being smaller than m), we see that the ith component in every ICV of every n-pc class is smaller than or at most equal to the ith component in every ICV of every m-pc class. Let us examine the comparison group #4/#8. Among the ICVs of the tetrad classes, the largest component to be found in indexes 1, 2, 4 and 5 is three, in index 3 four and in index 6 two. Among the ICVs of the octad classes, the smallest component to be found in indexes 1-5 is four, in index 6 two. In each of the 841 comparisons in the group, the ic contents of the tetrad class are completely embedded in those of the octad class, and the result is one of the uniform-valued value groups. From the point of view of SIM, the varying ic characteristics of the smaller classes have no significance whatsoever.

It seems natural to assume that cardinality difference is an important differentiating factor between SCs, but there is no reason to deem it so important that ic characteristics become totally meaningless. On the contrary, ic characteristics are the reason to reject SIM's suggestion that the degree of similarity between the SC pair {9-2,6-Z10} is identical to that between the pair {9-2,6-35}. The common SIM value is 21. %REL(9-2,6-Z10) = 5; %REL(9-2,6-35) = 56. Ex. 3.22.

EXAMPLE 3.22: Interval-class vectors of three SCs.

```
ICV(9-2)    = [7 7 7 6 6 3]
ICV(6-Z10)  = [3 3 3 3 2 1]
ICV(6-35)   = [0 6 0 6 0 3]
```

We saw in section 3.4.1.1 that SIM and the k measure are closely connected. As the numbers of distinct values are identical for all corresponding SIM and k value groups, it is obvious that the latter shares SIM's poor resolution in value groups #n/#m where there is a large difference between n and m. Also, it shares SIM's in-sensitivity to similarities between ic content profiles. All the triad-tetrad class pairs in Ex. 3.20, having a uniform SIM value, also have the uniform k value 3(3). Likewise, both nonad-hexad class pairs in Ex. 3.22 produce the k value 15(15).

SIM and k could be defended by saying that coarseness is not their property, but a property of the SC universe itself. In a sense this is true. All sufficiently small ic contents are embedded in all sufficiently large ones. In fact, if a measure observes extents of shared elements in ic contents and *does not* produce these uniform-valued value groups, it also observes something else, intentionally or otherwise. SIM and k process correctly what they were set to process, and we do not criticise them for showing the results they show, but for measuring what they measure. $\%REL_n$ functions better in terms of identifying grades of similarity between ic content profiles, but was previously seen as having problems of its own. (Section 3.4.2.1). It is in junctures like this that we identify the need for the total measures. Comparisons of limited materials seem always to produce at least some counterintuitive results, no matter what approach is selected.

3.6.3 Morris: ASIM

Absolute SIM. A modification of the SIM measure. Presented in Morris (1979-80). A similarity measure pairing interval-class instances by one-to-one correspondence.

COMPARISON PROCEDURE:

Absolute values of the differences between corresponding ICV components are added together. The sum is divided by the total number of ic instances in the two ICVs.

EQUATION:

$$\text{ASIM}(X,Y) = \frac{\text{SIM}(X,Y)}{\#\,\text{ICV}(X) + \#\,\text{ICV}(Y)}$$

EXAMPLE 3.23: ASIM(3-1,4-16). Prime forms and ICVs.[25]

```
3-1:  {0,1,2},   [2 1 0 0 0 0]
4-16: {0,1,5,7}, [1 1 0 1 2 1]
```

#ICV(3-1) + #ICV(4-16) = 3+6 = 9.

SIM(3-1,4-16) = 5. ASIM(3-1,4-16) = 5/9 ≈ 0.56.

EVALUATION CRITERIA FULFILLED:

C1, C2, C3.1, C3.2, C3.4, C4.

25 After Morris (1979-80:450).

THE ASIM VALUE GROUP #2-#12/#2-#12:

Non-integer values rounded to two decimal places. Value indicating highest degree of similarity: 0. Value indicating highest degree of dissimilarity: 1. Average: 0.42. Number of distinct values: 79. ASIM(X,Y) may or may not be ASIM(X_C,Y_C).

TABLE 3.5: The ASIM value groups #n/#m, 3 ≤ n,m ≤ 9. Lower values indicate higher degrees of similarity. Each table cell contains, clockwise from the top left: the lowest and highest values, the number of distinct values, the average.

	#3	#4	#5	#6	#7	#8	#9
#3	0.33 1.0 0.68 3						
#4	0.33 1.0 0.58 4	0.0 0.83 0.45 6					
#5	0.54 0.85 0.6 3	0.25 0.75 0.39 5	0.0 0.6 0.27 7				
#6	0.67 0.89 0.69 3	0.43 0.81 0.47 5	0.2 0.68 0.28 6	0.0 0.6 0.2 10			
#7	0.75 0.75 0.75 1	0.56 0.7 0.56 3	0.35 0.55 0.36 4	0.17 0.5 0.21 6	0.0 0.29 0.13 7		
#8	0.81 0.81 0.81 1	0.65 0.65 0.65 1	0.47 0.47 0.47 1	0.3 0.49 0.31 5	0.14 0.31 0.16 5	0.0 0.18 0.1 6	
#9	0.85 0.85 0.85 1	0.71 0.71 0.71 1	0.57 0.57 0.57 1	0.41 0.41 0.41 1	0.26 0.26 0.26 1	0.12 0.22 0.13 4	0.03 0.08 0.06 3

3.6.3.1 Analysis

SIM's most obvious problem, inability to meet the value commensurability criterion C3.1, is successfully solved in ASIM. As shown in Table 3.5, comparing values from different value groups is now meaningful. We see at a glance, for example, that the average values of the #n/#n value groups, 3 ≤ n ≤ 9, decrease rapidly as n grows. Or, that the ranges between minimum and maximum values are much wider in value groups involving two small cardinalities than in those involving two large ones.

Comparing the ASIM Table 3.5 to the SIM Table 3.4 shows that the numbers of distinct values in corresponding value groups are identical between SIM and ASIM. This means that ASIM too produces uniform-valued value groups. For example, the value group #6/#9 contains instances of the value 0.41 only. With the exception of the uniform scale of values, ASIM shares SIM's disadvantages. The reason is that the scaling of a SIM value with the vector component sums is done as the *last* step. (In %REL_n the vectors are scaled *before* comparison).[26] The poor resolution in

[26] When SCs of the same cardinality are compared, the order between scaling and comparing is not

some of the value groups remains, as does the insensitivity to similarities between the ic content profiles. Consequently, in Ex. 3.20 above, the ASIM values of the four triad-tetrad class comparisons would be uniform, 0.33. Both nonad-hexad class pairs in Ex. 3.22 produce the ASIM value 0.41, etc.

3.6.4 Lord: The Similarity Function (sf)

Presented in Lord (1981). A similarity measure pairing interval-class instances by one-to-one correspondence. For SCs of the same cardinality only.

COMPARISON PROCEDURE:

The sum of the absolute values of the differences between corresponding ICV components is divided by two.

EQUATION:

$$sf(X,Y) = \frac{\sum_{i=1}^{6} |x_i - y_i|}{2}$$

EXAMPLE 3.24: *sf*(6-1,6-Z4). Prime forms and ICVs.[27]

```
6-1:  {0,1,2,3,4,5},  [5 4 3 2 1 0]
6-Z4: {0,1,2,4,5,6},  [4 3 2 3 2 1]
```

sf(6-1,6-Z4) =

$$\frac{|5-4| + |4-3| + |3-2| + |2-3| + |1-2| + |0-1|}{2} =$$

$$\frac{1+1+1+1+1+1}{2} = 3$$

EVALUATION CRITERIA FULFILLED:

C3.3, C3.4, C4.

crucial. Because of this, there is an straight correlation between the entries in the #n/#n table cells of the %REL_2 Table 3.2 and the ASIM Table 3.5. The %REL_2 minimum, maximum and average entries are 100 times larger than their ASIM counterparts, with some minor differences due to rounding. The distinct value entries are identical. Between other value groups the correlation does not exist.

[27] After Lord (1981).

THE SET OF VALUES:[28]

All values are integers. Value indicating highest degree of similarity: 0. Value indicating highest degree of dissimilarity: 9. Average: 3. Number of distinct values: 10. *sf*(X,Y) = *sf*(X_C,Y_C).

TABLE 3.6: The *sf* value groups #n/#n, 3 ≤ n ≤ 9. Within a given value group, lower values indicate higher degrees of similarity. Each table cell contains: the lowest and highest values, the average, the number of distinct values.

#3/#3	1	3	2.05	3
#4/#4	0	5	2.69	6
#5/#5	0	6	2.7	7
#6/#6	0	9	3.05	10
#7/#7	0	6	2.7	7
#8/#8	0	5	2.69	6
#9/#9	1	3	2.05	3

3.6.4.1 Analysis

Lord's Similarity Function is almost identical to SIM, but was developed independently (Lord 1981:111). For every SC pair {X,Y} of the same cardinality, *sf*(X,Y) = SIM(X,Y)/2. According to Lord, the reason for halving the summation is to avoid counting each change in ic content twice - once where a given interval class is increased, and once for a corresponding decrease in another class (1981:93). Lord tested *sf* also with SCs of different cardinalities, but deemed the results disappointing. Probably he became aware of the problems we saw in SIM, a coarse degree of discrimination in some value groups and insensitivity to similarities between ic content profiles. A modification of *sf* which allows comparisons between SCs of different cardinalities is discussed, but no concrete results are provided (1981:109-10).

3.6.5 Isaacson: The IcVSIM Relation

The Interval-class Vector Similarity Relation. A scaled version of Teitelbaum's s.i. Presented in Isaacson (1990). Discussed at length also in Isaacson (1992). A similarity measure pairing interval-class instances by one-to-one correspondence.

28 The set of values comprises values in the nine value groups #n/#n, 2 ≤ n ≤ 10. As the measure does not meet the value commensurability criterion C3.1, this information is of limited importance.

COMPARISON PROCEDURE:

Corresponding ICV components are subtracted, the differences forming an *Interval-difference vector* (IdV).[29] The average of the IdV components is subtracted from each IdV component. The differences are squared and added together, the sum being divided by the length of the IdV, six. The final IcVSIM value is the square root of the quotient.[30]

EQUATION:

Given SCs X and Y and their vectors ICV(X) and ICV(Y), respectively, the Interval-difference vector is:

$$\text{IdV} = [(y_1-x_1)(y_2-x_2)\ldots(y_6-x_6)].$$

The value of the function IcVSIM(X,Y), then, is the degree of variance in the components of the IdV, measured with the standard deviation function σ. Defined in terms of the IdV, this function is:

$$\sigma = \sqrt{\frac{\sum\left(\text{IdV}_i - \overline{\text{IdV}}\right)^2}{6}}$$

where IdV_i is the ith component in the IdV and $\overline{\text{IdV}}$ the average of the components in the IdV (Isaacson 1990:18).

EXAMPLE 3.25: IcVSIM(6-1,3-1). Prime forms, ICVs and IdV.[31]

```
6-1: {0,1,2,3,4,5},  [5 4 3 2 1 0]
3-1: {0,1,2},        [2 1 0 0 0 0]
                 IdV [3 3 3 2 1 0]
```

IcVSIM(6-1,3-1)=

$$\sqrt{\frac{(3-2)^2+(3-2)^2+(3-2)^2+(2-2)^2+(1-2)^2+(0-2)^2}{6}}$$

$$=\sqrt{\frac{1+1+1+0+1+4}{6}}=\sqrt{\frac{8}{6}}=1.15$$

29 The order in which the two ICVs are taken does not affect the final result.

30 Isaacson's glossary entry for IcVSIM in (1992:255) is incorrect, suggesting that the square root of the sum of the squared differences is taken *before* the division.

31 The example is from Isaacson (1992:75-6).

EVALUATION CRITERIA FULFILLED:

C1, C2, C3.4.

THE ICVSIM VALUE GROUP #2-#12/#2-#12: [32]

Non-integer values rounded to two decimal places. Value indicating highest degree of similarity: 0. Value indicating highest degree of dissimilarity: 3.64. Average: 1.2. Number of distinct values: 121. If #X = #Y, IcVSIM(X,Y) = IcVSIM(X_C,Y_C). If #X ≠ #Y, IcVSIM(X,Y) may or may not be IcVSIM(X_C,Y_C).

TABLE 3.7: The IcVSIM value groups #n/#m, 3 ≤ n,m ≤ 9. Within a given value group, lower values indicate higher degrees of similarity. Each table cell contains, clockwise from the top left: the lowest and highest values, the number of distinct values, the average.

	#3	#4	#5	#6	#7	#8	#9
#3	0.58 1.53 0.93 7						
#4	0.5 2.14 0.99 13	0.0 2.24 1.12 15					
#5	0.37 2.19 1.03 15	0.47 2.75 1.11 19	0.0 2.45 1.13 19				
#6	0.0 3.06 1.14 21	0.5 3.55 1.2 28	0.37 3.29 1.18 29	0.0 3.46 1.25 31			
#7	0.58 2.24 1.17 14	0.5 2.81 1.23 20	0.37 2.54 1.17 19	0.0 3.32 1.2 29	0.0 2.45 1.13 19		
#8	0.37 2.11 1.28 14	0.47 2.49 1.32 19	0.0 2.77 1.24 21	0.37 3.58 1.24 29	0.37 2.73 1.12 20	0.0 2.24 1.12 15	
#9	0.76 2.14 1.42 13	0.58 2.71 1.44 20	0.47 2.49 1.34 18	0.5 3.25 1.3 22	0.5 2.29 1.12 15	0.47 2.29 1.03 15	0.58 1.53 0.93 7

3.6.5.1 Analysis

Although not initially conceived as such, IcVSIM turns out to be a scaled version of Teitelbaum's s.i. measure. In (1992:76) Isaacson defines also the latter in terms of the interval-difference vector. Ex. 3.26.

EXAMPLE 3.26: The s.i. measure defined in terms of the IdV.

$$\text{s.i.}(X,Y) = \sqrt{\sum_{i=1}^{6} \text{IdV}_i^2}$$

Instances of the lowest IcVSIM value, 0, are produced by Z-related SCs, as well as by

[32] As the measure does not meet the value commensurability criterion C3.1, this information is of limited importance.

a group of SC pairs with non-identical ICVs but with level IdVs.[33]

EXAMPLE 3.27: ICVs and IdV of the set-classes 6-30 and 3-10.

```
ICV(6-30)  = [2 2 4 2 2 3]
ICV(3-10)  = [0 0 2 0 0 1]
         IdV [2 2 2 2 2 2]
```

As the standard deviation function returns the value 0 for this interval-difference vector, IcVSIM(6-30,3-10) = 0.

The most important difference between IcVSIM and Teitelbaum's s.i. is that the former allows comparisons between SCs of different cardinalities, and the latter does not. Among the similarities is the shared inability to meet the crucial criteria C3.1 and C4 (calling for a a single scale of values for all comparisons and a uniform value for all comparable cases, respectively).

To examine the relation between IcVSIM and C4, let us compare again a number of SC pairs. Some of these were already compared with s.i. in Ex. 3.17.

EXAMPLE 3.28: Interval-class vectors of six triad classes.

```
ICV(3-1)   = [2 1 0 0 0 0]
ICV(3-10)  = [0 0 2 0 0 1]
ICV(3-11)  = [0 0 1 1 1 0]
ICV(3-12)  = [0 0 0 3 0 0]
ICV(3-9)   = [0 1 0 0 2 0]
ICV(3-7)   = [0 1 1 0 1 0]
```

The interval-class contents of the pairs {3-1,3-11}, {3-1,3-10} and {3-1,3-12} are disjoint. Despite this, the values for the three comparable pairs are different. IcVSIM(3-1,3-11) = 1.15; IcVSIM(3-1,3-10) = 1.29; IcVSIM(3-1,3-12) = 1.53. The reason for the varying values is the squaring of the IdV differences. IcVSIM rewards pairs with even distribution of unilaterally embedded ic instances, although the whole notion has no meaning from the point of view of the number of mutually embedded ones. (Section 2.4.3). Altogether, there are 25 SC pairs such that both classes are of cardinality 3 or larger and their ic contents are disjoint. To these pairs IcVSIM returns 10 different values, ranging from 1.15 to 2.14.

Furthermore, set-classes with partially shared ic contents with 3-1 may have as high IcVSIM values with it as SCs with disjoint ic contents. IcVSIM(3-1,3-9) = IcVSIM(3-1,3-11) = 1.15, even though k(3-1,3-9) = 1(2) and k(3-1,3-11) = 0(2). Here, as

[33] The latter pairs are given in Isaacson (1990:27-8 *n* 16).

with s.i., the reason is the component 2 in ICV(3-9). (Section 3.6.1.1). The pair {3-1,3-7}, in turn, is comparable with {3-1,3-9}, as in both the SCs share an instance of ic2. The two IcVSIM values are different, 1.0 and 1.15, respectively. The non-shared instances are distributed more evenly in ICV(3-7) than in ICV(3-9).

Another important feature of IcVSIM has to do with a notion we discussed in section 2.5: the status of a zero component in a vector. See Ex. 3.27 above, with the comparison IcVSIM(6-30,3-10). The two SCs were deemed maximally similar. From an arithmetical point of view it is correct to state that the distance from 2 to 0 is the same as the distance from 4 to 2 or from 3 to 1. From the point of view of the two ic contents, however, it is not so. Given the pairs of corresponding components 4,2 and 3,1, we compare something that exists to something that also exists. In the 2,0 case, by contrast, we compare something that exists to something that does not. Here only the *component* exists, not the corresponding ic. There is no "distance" from a non-existent ic instance to some number of existing instances. Interval-classes 1, 2, 4 and 5 do not contribute anything to 3-10, but they contribute more than half to the ic contents of 6-30. The two contents are highly different.[34] IcVSIM measures similarity between the *contours* of two ICVs (Rogers 1992).[35]

On the basis of the analysis above we must conclude that IcVSIM is an unreliable similarity measure. It is to be noted, however, that in cases where the ICVs contain only nonzero components, it can produce intuitively acceptable results. Ex. 3.29 gives the vectors of a maximally similar pair. IcVSIM(7-22,6-Z19) = 0.

EXAMPLE 3.29: ICVs and IdV of set-classes 7-22 and 6-Z19.

```
ICV(7-22)  = [4 2 4 5 4 2]
ICV(6-Z19) = [3 1 3 4 3 1]
         IdV [1 1 1 1 1 1]
```

The value corresponds meaningfully with the close distributional similarity between the two ICVs. %REL_2(7-22,6-Z19) = 6.

According to Isaacson, an important strength of IcVSIM is that it is able to show fine gradations of similarity, a property he thinks many other relations lack (Isaacson 1990:22). He offers no arguments to support the claim that a fine gradation

34 %REL_2(6-30,3-10) = 53.

35 In (1992:126-34) Isaacson gives an expanded version of IcVSIM, called EmbSIM. It is otherwise identical to IcVSIM, except for using 220-index *total-embedding vectors* (TeVs) that record all subset-class instances in a SC. As a great majority of components in TeVs of small-cardinality SCs are zeros, highly similar contours are guaranteed. The distortion that the flawed zero component status causes in IcVSIM is in much larger proportions in EmbSIM. According to Isaacson, EmbSIM is inadequate as a general measure of pcset similarity (Ibid., 131).

is an advantage. (Section 2.4.2.4).

3.6.6 Rogers: Distance Formula 1 ($IcVD_1$)

A modification of Morris's SIM measure. Presented in Rogers (1992). A similarity measure pairing interval-class instances by one-to-one correspondence.

COMPARISON PROCEDURE:

In both ICVs, each component is divided by the sum of all components in the vector. The absolute values of the differences between corresponding quotients are added together.

EQUATION:

$$IcVD_1(X,Y) = \sum_{i=1}^{6} \left| \frac{x_i}{\#\,ICV(X)} - \frac{y_i}{\#\,ICV(Y)} \right|$$

EXAMPLE 3.30: $IcVD_1$(3-1,4-19).[36] Prime forms, ICVs, sums of components and scaled interval-class vectors.

```
3-1:  {0,1,2},   [2 1 0 0 0 0], #ICV(3-1)  = 3, [2/3 1/3 0   0   0   0]
4-19: {0,1,4,8}, [1 0 1 3 1 0], #ICV(4-19) = 6, [1/6 0   1/6 3/6 1/6 0]
```

$IcVD_1$(3-1,4-19)= |2/3-1/6|+|1/3-0|+|0-1/6|+|0-3/6|+|0-1/6|+|0-0|

= 3/6 + 2/6 + 1/6 + 3/6 + 1/6 + 0 = 10/6 = 5/3 = 1.67.

EVALUATION CRITERIA FULFILLED:

C1, C2, C3.1, C3.2, C3.4, C4.

THE $ICVD_1$ VALUE GROUP #2-#12/#2-#12:

Non-integer values rounded to two decimal places. Value indicating highest degree of similarity: 0. Value indicating highest degree of dissimilarity: 2. Average: 0.59. Number of distinct values: 140. $IcVD_1(X,Y)$ may or may not be $IcVD_1(X_C,Y_C)$.

---

36 Examples 3.30, 3.31 and 3.33 are from Rogers (1992).

TABLE 3.8: The IcVD1 value groups #n/#m, 3 ≤ n,m ≤ 9. Lower values indicate higher degrees of similarity. Each table cell contains, clockwise from the top left: the lowest and highest values, the number of distinct values, the average.

	#3	#4	#5	#6	#7	#8	#9
#3	0.67 2.0 1.36 3						
#4	0.0 2.0 1.17 7	0.0 1.67 0.9 6					
#5	0.27 1.8 1.14 15	0.2 1.6 0.78 20	0.0 1.2 0.54 7				
#6	0.27 1.73 1.14 12	0.2 1.6 0.73 20	0.0 1.2 0.49 18	0.0 1.2 0.41 10			
#7	0.57 1.71 1.13 13	0.14 1.52 0.69 26	0.08 1.14 0.43 50	0.08 1.14 0.35 40	0.0 0.57 0.26 7		
#8	0.71 1.71 1.14 15	0.12 1.5 0.67 42	0.13 1.14 0.42 51	0.08 1.14 0.33 58	0.0 0.57 0.24 21	0.0 0.36 0.19 6	
#9	0.83 1.67 1.13 14	0.11 1.5 0.64 23	0.13 1.11 0.4 44	0.1 1.11 0.32 41	0.07 0.54 0.21 30	0.07 0.36 0.17 22	0.06 0.17 0.11 3

3.6.6.1 Analysis

IcVD1 is a similar modification of SIM as %REL2 is of *sf*. In both IcVD1 and %REL2, vector components are scaled *before* the absolute values of the differences are added together. In %REL2, each component is also multiplied by 100 after being divided by the component sum, and the sum of the absolute values is divided by 2. Consequently, %REL2(X,Y) = 50*IcVD1(X,Y). (Tables 3.2 and 3.8). Since %REL2 values are rounded, a value group in Table 3.2 may contain fewer distinct values than the corresponding value group in Table 3.8.

Due to the straight correlation between %REL2 and IcVD1, observations concerning the former are valid also for the latter. See sections 3.4.2 and 3.6.2.2.

3.6.7 Rogers: Distance Formula 2 (IcVD2)

A modification of Morris's SIM measure. Presented in Rogers (1992). A similarity measure pairing interval-class instances by one-to-one correspondence.

COMPARISON PROCEDURE:

In both ICVs, each component is divided by the square root of the sum of the squares of the components. Differences between corresponding quotients are squared and added together. The final result is the square root of the sum.

EQUATION:

$$\mathrm{IcVD2}(X,Y) = \sqrt{\sum \left(\frac{x_i}{\sqrt{\sum (x_i)^2}} - \frac{y_i}{\sqrt{\sum (y_i)^2}} \right)^2}$$

EXAMPLE 3.31: IcVD2(3-1,4-19). Prime forms and ICVs.

```
3-1: {0,1,2},     [2 1 0 0 0 0]
4-19:{0,1,4,8},  [1 0 1 3 1 0]
```

$$\sqrt{\sum(x_i)^2} = 2^2+1^2+0+0+0+0 = \sqrt{5}$$

$$\sqrt{\sum(y_i)^2} = 1^2+0+1^2+3^2+1^2+0 = \sqrt{12}$$

IcVD2(3-1,4-19) =

$$\sqrt{\left(\frac{2}{\sqrt{5}}-\frac{1}{\sqrt{12}}\right)^2+\left(\frac{1}{\sqrt{5}}-0\right)^2+\left(0-\frac{1}{\sqrt{12}}\right)^2+\left(0-\frac{3}{\sqrt{12}}\right)^2+\left(0-\frac{1}{\sqrt{12}}\right)^2+(0-0)^2}$$

$$= \sqrt{0.606^2+0.447^2+(-0.289)^2+(-0.866)^2+(-0.289)^2+0^2}$$

$$= \sqrt{1.483} = 1.218$$

EVALUATION CRITERIA FULFILLED:
C1, C2, C3.1, C3.2, C3.4.

THE ICVD2 VALUE GROUP #2-#12/#2-#12:
Non-integer values rounded to two decimal places. Value indicating highest degree of similarity: 0. Value indicating highest degree of dissimilarity: 1.41. Average: 0.54. Number of distinct values: 133. IcVD2(X,Y) may or may not be IcVD2(X_C,Y_C).

TABLE 3.9: The IcVD2 value groups #n/#m, $3 \le n,m \le 9$. Lower values indicate higher degrees of similarity. Each table cell contains, clockwise from the top left: the lowest and highest values, the number of distinct values, the average.

	#3	#4	#5	#6	#7	#8	#9
#3	0.67 1.41 1.11 10						
#4	0.0 1.41 0.95 42	0.0 1.33 0.83 52					
#5	0.27 1.3 0.89 58	0.19 1.3 0.71 76	0.0 1.09 0.59 59				
#6	0.27 1.3 0.87 63	0.19 1.3 0.67 90	0.0 1.09 0.52 87	0.0 1.0 0.45 63			
#7	0.43 1.17 0.85 52	0.14 1.16 0.64 79	0.1 0.94 0.46 70	0.08 0.94 0.38 75	0.0 0.64 0.31 37		
#8	0.53 1.16 0.84 44	0.16 1.08 0.62 74	0.14 0.94 0.44 66	0.08 0.94 0.35 72	0.0 0.64 0.27 50	0.0 0.46 0.23 24	
#9	0.63 1.1 0.84 29	0.16 1.06 0.62 58	0.18 0.87 0.43 51	0.12 0.87 0.33 53	0.09 0.53 0.24 35	0.07 0.41 0.19 31	0.09 0.25 0.15 7

3.6.7.1 Analysis

The (theoretical) maximum value is the same for all value groups, 1.41. As IcVD2 meets the value commensurability criterion C3.1, all values in Table 3.9 can be compared with one another without difficulty. IcVD2 has the same defect as s.i. and IcVSIM, however. It fails C4, the criterion calling for a uniform value for all comparable cases. The reason is also the same, i.e., the components are squared during the calculation.

Ex. 3.32 contains the ICVs of eight triad classes. The classes form six comparable pairs, given below. All have the uniform k value 1(2). Other factors differ: some ICVs are relatively peaked, others level; the corresponding nonzero components may be of the same or different sizes; etc.

EXAMPLE 3.32: Interval-class vectors of eight triad classes.

```
ICV(3-1)  = [2 1 0 0 0 0]
ICV(3-2)  = [1 1 1 0 0 0]
ICV(3-3)  = [1 0 1 1 0 0]
ICV(3-6)  = [0 2 0 1 0 0]
ICV(3-8)  = [0 1 0 1 0 1]
ICV(3-9)  = [0 1 0 0 2 0]
ICV(3-11) = [0 0 1 1 1 0]
ICV(3-12) = [0 0 0 3 0 0]
```

IcVD2(3-11,3-12) = 0.92; IcVD2(3-1,3-3) = 0.98; IcVD2(3-6,3-12) = 1.05; IcVD2(3-2,3-8) = 1.15; IcVD2(3-1,3-8) = 1.22; IcVD2(3-1,3-9) = 1.26. The value is different for every pair. The value group #3/#3 for IcVD 2 contains instances of 10 values. (Table 3.9). As mentioned earlier, from the viewpoint of mutually embedded ic instances there are only three categories available for the comparison group #3/#3.

3.6.8 Rogers: COSθ

Cosine Theta. Presented in Rogers (1992). A similarity measure pairing interval-class instances by one-to-many correspondence.

COMPARISON PROCEDURE:

The sum of the products of corresponding components is divided by the product of the square roots of the sums of the squares of the components.

EQUATION:

$$\cos\theta\,(X,Y) = \frac{\Sigma\, x_i * y_i}{\sqrt{\Sigma\,(x_i)^2} * \sqrt{\Sigma\,(y_i)^2}}$$

EXAMPLE 3.33: cosθ(3-1,4-19). Prime forms and ICVs.

```
3-1: {0,1,2},    [2 1 0 0 0 0]
4-19:{0,1,4,8},  [1 0 1 3 1 0]
```

$$\sqrt{\Sigma(x_i)^2} = 2^2+1^2+0\ +0\ +0\ +0\ = \sqrt{5}$$

$$\sqrt{\Sigma(y_i)^2} = 1^2+0\ +1^2+3^2+1^2+0\ = \sqrt{12}$$

$$\cos\theta(3\text{-}1,4\text{-}19) = \frac{[(2*1)+(1*0)+(0*1)+(0*3)+(0*1)+(0*0)]}{\sqrt{5} * \sqrt{12}} = \frac{2}{\sqrt{60}} = 0.258$$

EVALUATION CRITERIA FULFILLED:

C1, C2, C3.1, C3.2, C3.4.

THE COSθ VALUE GROUP #2-#12/#2-#12:

Non-integer values rounded to two decimal places. Value indicating highest degree of similarity: 1. Value indicating highest degree of dissimilarity: 0. Average: 0.81. Number of distinct values: 92. cosθ(X,Y) may or may not be cosθ(X_C,Y_C).

TABLE 3.10: The cos θ value groups #n/#m, 3 ≤ n,m ≤ 9. Higher values indicate higher degrees of similarity. Each table cell contains, clockwise from the top left: the lowest and highest values, the number of distinct values, the average.

	#3	#4	#5	#6	#7	#8	#9
#3	0.0 0.77 0.36 10						
#4	0.0 1.0 0.51 44	0.12 1.0 0.63 54					
#5	0.15 0.96 0.58 53	0.15 0.98 0.72 66	0.41 1.0 0.81 45				
#6	0.15 0.96 0.61 59	0.15 0.98 0.75 65	0.41 1.0 0.85 52	0.5 1.0 0.88 43			
#7	0.31 0.91 0.63 50	0.32 0.99 0.78 58	0.56 1.0 0.88 41	0.56 1.0 0.92 37	0.79 1.0 0.95 20		
#8	0.33 0.86 0.63 39	0.41 0.99 0.79 51	0.56 0.99 0.89 35	0.56 1.0 0.93 35	0.79 1.0 0.96 19	0.9 1.0 0.97 11	
#9	0.4 0.8 0.64 26	0.44 0.99 0.79 41	0.62 0.98 0.9 29	0.62 0.99 0.94 24	0.86 1.0 0.97 13	0.92 1.0 0.98 9	0.97 1.0 0.99 4

3.6.8.1 Analysis

Cosθ is the first one-to-many correspondence measure we examine. It was seen in section 3.2 that the one-to-one correspondence measures observe similarities between instance distributions. Obviously, taking the product of corresponding components does not answer distributionally oriented questions, like "how many mutually embedded ic instances are there?" A single product can result from many different component pairs. Rather, the values indicate "extents of pairedness" between two SCs, a notion not necessarily correlating with distributional similarity.

Ex. 3.34 gives four SC pairs containing the SC 3-11. We will first examine three of them, {3-11,3-2}, {3-11,3-9} and {3-11,3-12}.

EXAMPLE 3.34: Four pairs of ICVs belonging to triad classes.

```
ICV(3-11)  =  [0 0 1 1 1 0]        ICV(3-11)  =  [0 0 1 1 1 0]
ICV(3-2)   =  [1 1 1 0 0 0]        ICV(3-9)   =  [0 1 0 0 2 0]

ICV(3-11)  =  [0 0 1 1 1 0]        ICV(3-11)  =  [0 0 1 1 1 0]
ICV(3-12)  =  [0 0 0 3 0 0]        ICV(3-6)   =  [0 2 0 1 0 0]
```

Cosθ(3-11,3-2) = 0.33; cosθ(3-11,3-9) = 0.52; cosθ(3-11,3-12) = 0.58. In each case, only one pair of corresponding components is mutually nonzero, resulting in a product larger than zero. In the first case, one ic instance in 3-11 corresponds to one in 3-2. Then, one instance in 3-11 corresponds to two in 3-9 and, finally, one in 3-11 to three in 3-12. The values increase with the products, reflecting the fact that the number of instance pairings between the SCs increase. A one-to-one correspondence measure, by contrast, would produce a uniform value. The number of mutually embedded ic instances is the same in every case, one.

Cosθ does not meet C4, the criterion calling for a uniform value for all comparable cases. The divisor term $\sqrt{\Sigma(x_i)^2} * \sqrt{\Sigma(y_i)^2}$, needed to scale the values between 0 and 1, treats peaked and level vectors differently. Once again, the reason is in the squaring of the components. Let us examine the fourth set-class pair in Ex. 3.34, {3-11,3-6}. Cosθ(3-11,3-6) = 0.26, a value lower than 3-11 produced with 3-2. However, in both cases the product of the only pair of mutually nonzero components is the same, 1. ICV(3-6) contains a larger component than ICV(3-2), resulting in a larger divisor and a lower value. Due to the relation with C4, we will not use cosθ to analyse the differences between one-to-one and one-to-many correspondence measures.

3.6.9 Rahn: The Ak Measure

Absolute (Adjusted) k measure. A modification of Morris's k measure (k number). Presented in Rahn (1979-80). A similarity measure pairing interval-class instances by one-to-one correspondence.[37]

COMPARISON PROCEDURE:
The smaller components in each pair of corresponding ICV components are added together. The sum is multiplied by two and divided by the total number of ic instances in the two ICVs.

EQUATION:

$$ak(X,Y) = \frac{2 * k(X,Y)}{\# ICV(X) + \# ICV(Y)}$$

EXAMPLE 3.35: ak(5-1,5-16). Prime forms and ICVs.[38]

```
5-1:  {0,1,2,3,4}, [4 3 2 1 0 0]
5-16: {0,1,3,4,7}, [2 1 3 2 1 1]
```

2 * k(5-1,5-16) = 12. #ICV(5-1) + #ICV(5-16) = 20. ak(5-1,5-16) = 12/20 = 0.6.

---

[37] Rahn uses the name ak(A,B). The name ak measure is adopted here.

[38] The classes in Ex. 3.35 are the same as in the k measure example 3.1. From Rahn (1979-80).

EVALUATION CRITERIA FULFILLED:

C1, C2, C3.1, C3.2, C3.4, C4.

THE AK VALUE GROUP #2-#12/#2-#12:

Non-integer values rounded to two decimal places. Value indicating highest degree of similarity: 1. Value indicating highest degree of dissimilarity: 0. Average: 0.58. Number of distinct values: 78. ak(X,Y) may or may not be ak(X_C,Y_C).

TABLE 3.11: The ak measure value groups #n/#m, 3 ≤ n,m ≤ 9. Higher values indicate higher degrees of similarity. Each table cell contains, clockwise from the top left: the lowest and highest values, the number of distinct values, the average.

	#3	#4	#5	#6	#7	#8	#9
#3	0.0 0.67 0.32 3						
#4	0.0 0.67 0.42 4	0.17 1.0 0.55 6					
#5	0.15 0.46 0.4 3	0.25 0.75 0.61 5	0.4 1.0 0.73 7				
#6	0.11 0.33 0.31 3	0.19 0.57 0.53 5	0.32 0.8 0.72 6	0.4 1.0 0.8 10			
#7	0.25 0.25 0.25 1	0.3 0.44 0.44 3	0.45 0.65 0.64 4	0.5 0.83 0.79 6	0.71 1.0 0.87 7		
#8	0.19 0.19 0.19 1	0.35 0.35 0.35 1	0.53 0.53 0.53 1	0.51 0.7 0.69 5	0.69 0.86 0.84 5	0.82 1.0 0.9 6	
#9	0.15 0.15 0.15 1	0.29 0.29 0.29 1	0.43 0.43 0.43 1	0.59 0.59 0.59 1	0.74 0.74 0.74 1	0.78 0.87 0.86 4	0.92 0.97 0.94 3

3.6.9.1 Analysis

The correlation between k and SIM values was given formally in section 3.4.1.1 above. Predictably, there is also a correlation between the values produced by the adjusted versions ak and ASIM. For every SC pair {X,Y},

(1) ASIM(X,Y) = (1 - ak(X,Y)).
(2) ak(X,Y) = (1 - ASIM(X,Y)).[39]

The correlation becomes evident from comparing the ASIM value group information table 3.5 to the ak Table 3.11. The numbers of distinct values are identical between corresponding value groups. Also, the sum of corresponding average values is 1, and the sum of a minimum value in one table and the corresponding maximum

39 Isaacson (1992:45).

value in the other is also 1, with a few small exceptions due to rounding. It is also because of rounding that the total numbers of distinct values differ slightly between the two measures, being 79 for ASIM and 78 for ak.

The ak measure has an advantage over k in that it provides a single scale of values for all comparisons. However, it also shares the disadvantages of k, SIM and ASIM: a poor degree of discrimination in many comparison groups, as well as insensitivity to similarity between ic content profiles. (Section 3.6.2.2). Normalising k values into ak values is done at the same stage as the normalisation of SIM values into ASIM values, as the last step. Consequently, the values are scaled, but the other counterintuitive features remain. Comparisons between tables 3.1, 3.4, 3.5 and 3.11 show that the uniform-valued value groups are exactly the same for k, SIM, ASIM and ak.

3.6.10 Rahn: $MEMB_n$

Mutual Embedding. Presented in Rahn (1979-80). A similarity measure pairing subset-class instances by one-to-many correspondence, one subset-class cardinality at a time.

COMPARISON PROCEDURE:

Corresponding n-class vector components are added together if both are nonzero. The final value is the sum of all such component pair sums.

EQUATION:[40]

Given SCs X and Y, cardinality-class n, #X,#Y $\geq$ n, and the value of the function EMB(A,X), being the number of instances of SC A in X,

$$MEMB_n(X,Y) = EMB(A,X) + EMB(A,Y)$$

for all A in n so that EMB(A,X) > 0 and EMB(A,Y) > 0.

[40] Rahn uses the expression $MEMB_n$(A,X,Y). We will use the shorter expression $MEMB_n$(X,Y) to have a uniform notation for all measures and argument SCs.

EXAMPLE 3.36: MEMB2(4-17,5-7). Prime forms and 2CVs.[41]

```
4-17: {0,3,4,7},         [1 0 2 2 1 0]
5-7:  {0,1,2,6,7},       [3 1 0 1 3 2]
MEMB2(4-17,5-7)     =    4  +  3+4     = 11
```

EVALUATION CRITERIA FULFILLED:

C1, C2, C3.3, C3.4, C4.[42]

THE MEMB2 VALUE GROUP #2-#12/#2-#12:[43]

All values are integers. Value indicating highest degree of similarity: 121. Value indicating highest degree of dissimilarity: 0. Average: 30. Number of distinct values: 79. MEMB2(X,Y) may or may not be MEMB2(X_C,Y_C).

TABLE 3.12: The MEMB2 value groups #n/#m, 3 ≤ n,m ≤ 9. Within a given value group, higher values indicate higher degrees of similarity. Each table cell contains, clockwise from the top left: the lowest and highest values, the number of distinct values, the average.

	#3	#4	#5	#6	#7	#8	#9
#3	0 5 2.32 5						
#4	0 9 4.72 9	3 12 8.38 10					
#5	3 13 7.11 11	4 16 12.65 12	8 20 18.44 11				
#6	4 18 9.39 13	5 21 16.49 16	10 25 23.57 13	12 30 29.16 11			
#7	6 18 12.08 13	11 27 20.95 17	19 31 29.4 12	24 36 35.43 11	42 42 42.0 1		
#8	7 21 15.07 15	13 34 25.91 20	22 38 35.84 13	27 43 42.23 13	49 49 49.0 1	56 56 56.0 1	
#9	9 24 18.56 14	15 42 31.63 22	26 46 43.26 11	31 51 50.02 11	57 57 57.0 1	64 64 64.0 1	72 72 72.0 1

41 After Isaacson (1990).

42 The relation between C4 and $MEMB_n$ differs from that between C4 and the measures based on one-to-one correspondence. A uniform number of shared ic instances, our usual requirement, is inapplicable since the formula calls for observing something else than this. $MEMB_n$ is its own criteria for C4. Every SC pair in a comparison group with a uniform sum of mutually nonzero component pair sums gets a uniform value. Other vector characteristics do not affect the result.

43 As the measure does not meet the value commensurability criterion C3.1, this information is of limited importance.

3.6.10.1 Analysis

Rahn does not discuss the status of $MEMB_n$ in (1979-80), but it is probably just a preliminary version of the total measure TMEMB, presented only to describe the basic approach in a simple manner (Isaacson 1990:27 *n* 8). TMEMB, in turn, leads to ATMEMB, the version normalising TMEMB values between 0 and 1. That ATMEMB is Rahn's final goal seems evident from the fact that both $MEMB_n$ and TMEMB fail to produce a uniform scale of values for all value groups, a defect Rahn identifies and wants to avoid (Rahn 1979-80:493). The inability to meet C3.1 is corrected in ATMEMB.

Let us compare again the same four SC pairs as in section 3.6.2.2 above, namely 3-5 with each of the SCs 4-9, 4-6, 4-5 and 4-Z15. The 2CVs are given in Ex. 3.37.

EXAMPLE 3.37: The 2-class vectors of five SCs.

```
2CV(3-5)   =  [1 0 0 0 1 1]
2CV(4-9)   =  [2 0 0 0 2 2]
2CV(4-6)   =  [2 1 0 0 2 1]
2CV(4-5)   =  [2 1 0 1 1 1]
2CV(4-Z15) =  [1 1 1 1 1 1]
```

$MEMB_2$(3-5,4-9) = 9; $MEMB_2$(3-5,4-6) = 8; $MEMB_2$(3-5,4-5) = 7; $MEMB_2$(3-5,4-Z15) = 6. The decreasing intuitive similarity between the SCs in the four pairs corresponds with the decreasing values. In the ICVs, the number of component pairs containing both a zero and a nonzero element increases, and these pairs do not contribute to the outcome. The problem of SIM, i.e., insensitivity to the similarity between ic content profiles, is avoided here.

This does not indicate, however, that the problem would be avoided in every case. When the n-class vectors of SCs X and Y contain only nonzero components, all pairs of corresponding components fulfil the condition of containing only nonzero elements and contribute to the $MEMB_n$ value. In these cases, $MEMB_n$(X,Y) = #nCV(X) + #nCV(Y). For example, as all 2CVs belonging to SCs of cardinality 7 or higher contain only nonzero components (Forte 1973a:20), all comparison groups #n/#m such that n, m $\geq$ 7 will produce uniform-valued $MEMB_2$ value groups. (Table 3.12). In these cases the ic characteristics of the compared SCs lose their meaning entirely. And even if a given value group contains instances of more than one value, the value distribution may suggest a very poor resolution. In the $MEMB_2$ value group #6/#6, for example, almost three quarters of the values are instances of

30, the value indicating maximal similarity. The average value is 29.16.

3.6.10.2 MEMB$_n$ and the One-to-One Correspondence Measures

In section 3.2 we examined the notion of one-to-many correspondence with an example observing the number of pairings. In MEMB$_n$, the viewpoint to one-to-many correspondence is slightly different. Suppose we have some SCs X and Y, containing one and five instances of some subset-class S, respectively. The sum, six, is distant from one of the components, and could have been produced by component pairs {2,4} and {3,3} as well. A single sum, like a single product, can reflect many degrees of distributional (one-to-one correspondence) similarity. When the sole instance of S in X is paired with each of the five instances in Y, the sum is not an answer to the question "how many pairings are there?" The answer to that is 5, requiring multiplication. Rather, the question is, "how many instances of S participate in a pairing?"

In MEMB$_n$, each sum of mutually nonzero components is twice the arithmetic mean (average) of the components (Lewin 1979-80b:500). When using the one-to-one correspondence measures, we focus on component size differences. In MEMB$_n$, we pool instances belonging to mutually represented subset-classes together.

Let us examine how MEMB$_n$ fares with the one-to-one correspondence measures. We shall first compare the dyad class contents of two pairs, {3-3,3-8} and {3-3,3-12}. The prime forms, 2CVs and 2C%Vs of the classes are given in Ex. 3.38 below.

EXAMPLE 3.38: Prime forms, 2CVs, and 2C%Vs of three SCs

```
3-3:  {0,1,4}  [1 0 1 1 0 0] [33 0 33  33 0  0]
3-8:  {0,2,6}  [0 1 0 1 0 1] [0 33  0  33 0 33]
3-12: {0,4,8}  [0 0 0 3 0 0] [0  0  0 100 0  0]
```

First, let us set our focus on one-to-one correspondence, and, consequently, on similarities between instance distributions. We see that the dyad class contents of the SCs are rather different, but have one aspect in common. All contain at least one instance of the dyad class 2-4. When we pair the instances by one-to-one correspondence, we see that some of them remain without counterparts in both SC pairs.

In the pair {3-3,3-8}, the non-paired instances belong to 2-1 and 2-3 in the former triad class, to 2-2 and 2-6 in the latter. The first dyad class pair is not contained at all in the second triad class, and vice versa. In the pair {3-3,3-12}, the situation is similar with 3-3 but different with 3-12. In the latter, the two non-paired instances belong to the dyad class 2-4 that is represented also in 3-3.

From the point of view of one-to-one correspondence, it does not matter that the non-paired instances in a sense represent two different categories. The number of shared instances is one in both cases. Consequently, if we compare the pairs with SIM and %REL2, the measures produce uniform values. SIM(3-3,3-8) = SIM(3-3,3-12) = 4. %REL2(3-3,3-8) = %REL2(3-3,3-12) = 67.

If we then compare the same pairs with MEMB2, the values are different and suggest closer similarity to the latter pair. MEMB2(3-3,3-8) = 2, MEMB2(3-3,3-12) = 4. The results seem intuitively acceptable. The number of 2-4 instances participating in a pairing is higher in the latter pair. From the one-to-one correspondence viewpoint we would say that in {3-3,3-8}, the non-paired instances are completely unrelated, whereas in {3-3,3-12} half of them, the two 2-4 instances in 3-12, are related to the paired instances. The entire dyad class contents of 3-12 contribute to the similarity between it and 3-3. Here, in terms of correlating measured values with intuitive similarity, it seems that MEMB2 has an edge over the one-to-one correspondence measures.

EXAMPLE 3.39: Prime forms, 2CVs and 2C%Vs of four SCs

```
4-1:  {0,1,2,3},  [3 2 1 0 0 0],  [50 33 17 0 0  0]
4-28: {0,3,6,9},  [0 0 4 0 0 2],  [ 0  0 67 0 0 33]

4-3:  {0,1,3,4},  [2 1 2 1 0 0],  [33 17 33 17  0  0]
4-9:  {0,1,6,7},  [2 0 0 0 2 2],  [33  0  0  0 33 33]
```

Ex. 3.39 gives two more pairs, {4-1,4-28} and {4-3,4-9}. SIM and %REL2 deem 4-1 and 4-28 highly dissimilar. The two values, SIM(4-1,4-28) = 10 and %REL2(4-1,4-28) = 83, are instances of the highest values in their value groups. Both measures suggest closer similarity to the latter pair. SIM(4-3,4-9) = 8, %REL2(4-3,4-9) = 67. 4-1 and 4-28 have one instance of 2-3 in common, 4-3 and 4-9 two instances of 2-1.

According to MEMB2, the first pair represents a higher degree of similarity than the second. MEMB2(4-1,4-28) = 5, MEMB2(4-3,4-9) = 4. The number of paired dyad class instances is higher in the former case than in the latter.

Unlike the case in Ex. 3.38, MEMB2 seems now to produce values contradicting the intuitive sense of similarity. The dyad class contents of 4-28 contain instances of only two classes, one of them, 2-3, having a stronger representation than any other dyad class in any other 4-pc class. This exceptional concentration is in all likelihood a factor *distancing* 4-28 from other classes, not one bringing it closer to them. 4-3 and 4-9 do not share a majority of their dyad class contents either, but the extent of similarity they have seems to result from the fact that the two instances of 2-1 in the former are matched by as many in the latter.

We could examine many more SC pairs besides those in examples 3.38 and 3.39, but our point is already evident. One-to-one correspondence measures are insensitive to non-paired instances of subset-classes that are represented in both SCs. $MEMB_2$, in turn, is insensitive to similarities between instance distributions. The strength of the one approach is precisely the weakness of the other.

3.7 THE TOTAL MEASURES

3.7.1 Rahn: TMEMB

Total Mutual Embedding. An expanded version of the $MEMB_n$ measure. Presented in Rahn (1979-80). A total measure pairing subset-class instances by one-to-many correspondence.

COMPARISON PROCEDURE:
TMEMB(X,Y) is the sum of all $MEMB_n$(X,Y) values, n ranging from 2 to the lesser of #X, #Y.

EQUATION:[44]

$$TMEMB(X,Y) = \sum_{n=2}^{12} MEMB_n(X, Y)$$

EXAMPLE 3.40: TMEMB(5-7,5-15). The prime forms are {0,1,2,6,7} and {0,1,2,6,8}, respectively. As the cardinality of both SCs is five, the value of TMEMB(5-7,5-15) is $MEMB_2$(5-7,5-15) + $MEMB_3$(5-7,5-15) + $MEMB_4$(5-7,5-15) + $MEMB_5$(5-7,5-15). Both 5CV(5-7) and 5CV(5-15) will contain only one nonzero component, 1, in the index corresponding to the class itself. Sums of component pairs containing zeros are not shown.

```
    2CV(5-7)   = [3 1 0 1 3 2]           3CV(5-7)   = [1 0 0 2 5 0 0 1 1 0 0 0]
    2CV(5-15)  = [2 2 0 2 2 2]           3CV(5-15)  = [1 0 0 2 2 0 0 4 1 0 0 0]
MEMB2(5-7,5-15)= 5+3 + 3+5+4=20; MEMB3(5-7,5-15)  =  2  +  4+7  +  5+2       =20

   4CV(5-7)   = [0 0 0 0 1 1 0 1 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0]
   4CV(5-15)  = [0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 1 0 0 0 0]
MEMB4(5-7,5-15)=        3           +           3                          =6
```

[44] The equation, given in Rahn (1979-80:492), sets the upper subset-class cardinality limit at 12. According to Rahn this is harmless, as there are no mutually embedded subset-classes of cardinalities greater than the lesser of #X and #Y.

```
5CV(5-7) =[0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
5CV(5-15)=[0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
MEMB5(5-7,5-15)                                                                      =0
```

TMEMB(5-7,5-15) = 20 + 20 + 6 + 0 = 46.

EVALUATION CRITERIA FULFILLED:

C1, C2, C3.3, C4, C5.

THE TMEMB VALUE GROUP #2-#12/#2-#12:[45]

All values are integers. Value indicating highest degree of similarity: 6,118. Value indicating highest degree of dissimilarity: 0. Average: 131. Number of distinct values: 877. TMEMB(X,Y) may or may not be TMEMB(X_C,Y_C).

TABLE 3.13: The TMEMB value groups #n/#m, 3 ≤ n,m ≤ 9. Within a given value group, higher values indicate higher degrees of similarity. Each table cell contains, clockwise from the top left: the lowest and highest values, the number of distinct values, the average.

	#3		#4		#5		#6		#7		#8		#9	
#3	0	5												
	2.32	5												
#4	0	14	3	19										
	5.3	13	10.38	17										
#5	3	20	4	33	10	46								
	8.49	15	18.1	29	31.68	33								
#6	4	31	5	52	13	82	16	103						
	11.81	19	26.15	40	48.59	58	74.84	66						
#7	6	31	16	59	32	102	45	164	112	228				
	15.93	20	36.35	42	70.74	62	113.97	96	178.13	80				
#8	7	32	19	79	39	152	52	255	173	366	266	468		
	20.71	22	48.57	50	98.1	81	162.15	124	263.81	136	389.39	116		
#9	11	37	22	95	63	182	93	309	252	510	325	741	739	980
	26.56	17	64.12	56	134.93	76	229.74	116	389.07	117	589.75	135	897.76	38

3.7.1.1 Analysis

According to Rahn, $MEMB_n$(X,Y) is naturally generalised to a measure of the total mutual embedding in X and Y of subset-classes of all sizes greater than one (1979-80:493). In cases where #X = #Y but X ≠ Y, the #X-class vectors do not contribute anything to the TMEMB value. Like $MEMB_n$, TMEMB also produces different scales of values for different value groups. (Table 3.13). Now that a large number of subset-

[45] As the measure does not meet the value commensurability criterion C3.1, this information is of limited importance.

class instances of different cardinalities participate in the comparisons, the scales vary to the extent that identification of comparable degrees of similarity, even approximately, may be very complicated. The extreme values in the value group #3/#3, for example, are 0 and 5. In the group #5/#6 they are 13 and 82, in the group #9/#9 739 and 980, etc.

There are no uniform-valued value groups, meaning that TMEMB can discriminate between cases which $MEMB_2$ deemed equally similar. For example, it was mentioned above that almost three quarters of the values in the $MEMB_2$ value group #6/#6 are instances of the maximum value 30. When these uniform-valued cases are compared separately with TMEMB, the resulting set of values ranges from 55 to 103, containing instances of 45 distinct values.

As TMEMB and ATMEMB are expanded versions of $MEMB_n$, we do not examine them separately as one-to-many correspondence measures. The observations made in section 3.6.10.2 are valid also with respect to the two total measures.

3.7.1.2 TMEMB and Criteria C5 and C6

TMEMB is the first total measure we examine. It can discriminate between Z-related classes and, under T_n-classification, between inversionally related classes. In the following we will analyse aspects of the measure itself, not the values it produces. We saw above that the latter are difficult to interpret due to TMEMB's inability to meet the value commensurability criterion C3.1. The value analysis will be offered in connection with the adjusted version ATMEMB.

Our present interests are connected to the discussion in section 2.5 above, concerning the status of different subset-class cardinalities. We will use the comparison group #9/#9 as our example, obtaining its $MEMB_8$ and TMEMB value groups. As the value of each #9/#9 comparison $MEMB_8$(X,Y) is both an independent result and a part of the corresponding TMEMB(X,Y) value, we compare the two value groups in order to assess what sort of impact the $MEMB_8$ values have in their TMEMB counterparts. Also, we analyse what the results mean from the point of view of the non-common subset-class criterion C6.

Among the 66 values in the $MEMB_8$ value group #9/#9, the lowest and highest values are 0 and 8, respectively. The average value is slightly less than 4. Ex. 3.41 gives the 8CVs of two SC pairs producing instances of the extreme values. $MEMB_8$(9-3,9-9) produces the minimum value, $MEMB_8$(9-5,9-8) the maximum.

EXAMPLE 3.41.a: MEMB8(9-3,9-9) = 0. The prime forms are {0,1,2,3,4,5,6,8,9} and {0,1,2,3,5,6,7,8,10}, respectively.

```
8CV(9-3)       = [0 1 1 1 0 0 1 0 0 0 0 1 0 0 1 0 1 1 1 0 0 0 0 0 0 0 0 0 0]
8CV(9-9)       = [0 0 0 0 0 1 0 0 0 0 0 0 0 2 0 2 0 0 0 0 0 2 2 0 0 0 0 0 0]
MEMB8(9-3,9-9) =                                                          =0
```

EXAMPLE 3.41.b: MEMB8(9-5,9-8) = 8. The prime forms are {0,1,2,3,4,6,7,8,9} and {0,1,2,3,4,6,7,8,10}, respectively.

```
    8CV(9-5) =[0 0 0 0 1 1 0 1 1 0 0 0 1 0 1 1 0 1 0 0 0 0 0 0 0 0 0 0 1]
    8CV(9-8) =[0 0 0 0 1 0 0 0 0 0 0 1 0 0 1 1 0 0 0 0 1 0 0 1 1 0 1 0 1]
MEMB8(9-5,9-8)=        2           +        2+2            +             2=8
```

In the TMEMB value group #9/#9, in turn, the lowest and highest values are 739 and 980, respectively. The average is approximately 898. The contribution of the 8-pc subset-classes to the whole TMEMB outcome is very small indeed. Given all nonad class pairs {X,Y}, there is not a single case where the MEMB8(X,Y) value would be a full one percent of the corresponding TMEMB(X,Y) value.

We can draw two very different conclusions from this observation. On the one hand, if we set out to count instances of mutually embedded octad classes, this is exactly the outcome we will get. The minuscule octad class representations reflect the subset-class properties of the nonad classes, not the properties of TMEMB. On the other hand, if we set out to compare nonad classes with the help of the octad classes included in them, we know that in a space of only 12 pitch-classes there is simply no room for an 8-element object to be *completely* dissimilar from another 8-element object. They have to share something, and that something can be identified and utilized when assessing the degree of similarity between them. Consequently, two collections of octad classes may be disjoint, but still each element in one has much more to do with the elements in the other than, say, two elements have in two disjoint collections of dyad classes. That the "way of being disjoint" is not uniform for small and large-cardinality subset-class contents is, of course, the very observation that gave rise to C6.

In Ex. 3.41.b, the mutually embedded octad classes are a meaningful indication of similarity between 9-5 and 9-8, but not the only one the octad classes can provide. Conversely, in Ex. 3.41.a, the fact that 9-3 and 9-9 do not share any octad classes does not mean that the octad class contents contribute to dissimilarity only. For example, 8CV(9-3) contains an instance of 8-Z15 and 8CV(9-9) two instances of 8-16. If we compare these two octad classes separately with TMEMB, we find that among the 406 values in the TMEMB value group #8/#8, the resulting value, 447, is in the highest 4% of values. The maximum is 468, average 389. There are other cross-correlated octad class pairs getting almost as high values. For example, TMEMB(8-2,8-22) = 444, TMEMB(8-4,8-14) = 440, etc. Even such a superficial glance into the disjoint octad class contents reveals obvious similarities.

3.7.2 Rahn: ATMEMB

Absolute (Adjusted) Total Mutual Embedding. A modification of the TMEMB measure. Presented in Rahn (1979-80). A total measure pairing subset-class instances by one-to-many correspondence.

COMPARISON PROCEDURE:

Given SCs X and Y, the value of TMEMB(X,Y) is first taken. The number of all subset-class instances of cardinality 2 or larger in X is found out and added to the number of all subset-class instances of cardinality 2 or larger in Y. The TMEMB(X,Y) value is divided by the sum of these subset-class instances.

EQUATION:

$$\text{ATMEMB}(X,Y) = \frac{\text{TMEMB}(X,Y)}{2^{\#X} + 2^{\#Y} - (\#X + \#Y + 2)}$$

EXAMPLE 3.42: ATMEMB(5-7,5-15). The TMEMB value of these SCs, 46, was found out in Ex. 3.40 above. The example also gave the prime forms and the vector pairs. The value of the divisor term in the ATMEMB equation is $(2^5 + 2^5 - (5 + 5 + 2)) = (32 + 32 - 12) = 52$. The ATMEMB value is $46/52 \approx 0.88$. It is an instance of the highest value in the ATMEMB value group #5/#5.

EVALUATION CRITERIA FULFILLED:

C1, C2, C3.1, C3.2, C3.4, C4, C5.

THE ATMEMB VALUE GROUP #2-#12/#2-#12:

Non-integer values rounded to two decimal places. Value indicating highest degree of similarity: 1. Value indicating highest degree of dissimilarity: 0. Average: 0.45. Number of distinct values: 101. ATMEMB(X,Y) may or may not be ATMEMB(X_C,Y_C).

TABLE 3.14: ATMEMB value groups #n/#m, 3 ≤ n,m ≤ 9. Higher values indicate higher degrees of similarity. Each table cell contains, clockwise from the top left: the lowest and highest values, the number of distinct values, the average.

	#3	#4	#5	#6	#7	#8	#9
#3	0.0 0.62 0.29 5						
#4	0.0 0.93 0.35 13	0.14 0.86 0.47 17					
#5	0.1 0.67 0.28 15	0.11 0.89 0.49 29	0.19 0.88 0.61 33				
#6	0.07 0.51 0.19 19	0.07 0.76 0.38 40	0.16 0.99 0.59 58	0.14 0.9 0.66 59			
#7	0.05 0.25 0.13 18	0.12 0.45 0.28 32	0.22 0.7 0.48 45	0.25 0.93 0.64 61	0.47 0.95 0.74 39		
#8	0.03 0.13 0.08 11	0.07 0.31 0.19 23	0.14 0.56 0.36 35	0.17 0.84 0.53 54	0.47 1.0 0.72 48	0.54 0.95 0.79 38	
#9	0.02 0.07 0.05 6	0.04 0.19 0.13 16	0.12 0.34 0.26 21	0.17 0.55 0.41 33	0.41 0.82 0.63 32	0.43 0.99 0.79 38	0.74 0.98 0.89 18

3.7.2.1 Analysis

A SC of cardinality n contains 2^n subset-class instances. Out of these, the n instances of the monadic class 1-1, as well as the single instance of the null set-class 0-1, are of no interest from the point of view of subset-class contents. In the divisor of the ATMEMB equation, the sum (#X + #Y + 2) equals the number of these instances. It is subtracted from the sum $2^{\#X} + 2^{\#Y}$, the difference giving the number of subset-class instances of cardinality 2 and larger. Thus, ATMEMB returns the ratio between the number of instances belonging to mutually embedded subset-classes of cardinality 2 or larger and the number of all subset-class instances of cardinality 2 or larger.

ATMEMB is related to TMEMB in the same way as ASIM is related to SIM and ak to k. All unadjusted values are scaled between 0 and 1. The advantage ATMEMB has over TMEMB is obvious. Since the scale of values is now uniform for all cases, comparisons can be made without difficulty. Table 3.14 shows, for example, that value group #n/#n averages grow rather rapidly as n grows. The highest value in #3/#3, 0.62, is lower than the lowest value, 0.74, in #9/#9.

The value of the divisor term in the ATMEMB equation is the same for all SC pairs in a single comparison group. Consequently, given SC pairs {X,Y} and {Z,W} of the same comparison group, if TMEMB(X,Y) = TMEMB(Z,W), also ATMEMB(X,Y) = ATMEMB(Z,W). This seems to imply that TMEMB and ATMEMB value groups produced by a given comparison group always contain the same number of distinct values. This is not true in all cases, however. Rounding the ATMEMB values to two decimal places may lead to the result that several TMEMB values correspond to only one ATMEMB value. Because of this, some distinct value entries in the ATMEMB Table 3.14 are lower than their counterparts in the TMEMB Table 3.13.

The divisor in the ATMEMB equation creates an interesting dilemma. Rahn states that the divisor gives "the total number of possible subsets of size greater than one that are mutually embeddable in sets A and B." (1979-80:493). This, however, is true only if #A = #B. Let X be a triad class and Y a nonad class. During TMEMB calculation we would compare 2CV(X) to 2CV(Y) and 3CV(X) to 3CV(Y). No other n-class vectors would participate as the limit is the lesser of #X and #Y (Ibid., 493). When determining the divisor to obtain the ATMEMB(X,Y) value, however, *all* subset-class instances of cardinality 2 or larger would be counted. For X, the number of the instances is $(2^3 - 3 - 1) = 4$. For Y, the corresponding figure is $(2^9 - 9 - 1) = 502$. Y's 4-pc and larger subset-classes are not "mutually embeddable" as they are of larger cardinalities than X. Out of the 502 subset-class instances in Y, only 120 dyad class and triad class instances fulfil the condition. The remaining 382 instances increase the size of the divisor considerably, and, consequently, considerably decrease the degree of similarity between X and Y.

It would seem natural to assume that if #Y > #X, only those subset-class instances in Y whose cardinality is up to #X could contribute to the divisor. This would create a natural balance between the number of instances belonging to mutually embedded subset-classes (the TMEMB value), and the number of instances belonging to subset-classes that in principle *could* be mutually embedded (constituting the divisor).[46] One can only guess whether the participation of the mutually non-embeddable subset-class instances in the larger SC is a mistake or an intentionally adopted feature. Rahn may have come to the conclusion that taking all instances into account correlates better with the intuitive sense of similarity he experiences. In this case, the statement about the mutually embeddable instances is incorrect. No examples are provided in Rahn (1979-80). Nor is this detail clarified or any examples provided in Rahn (1989), another source where ATMEMB is discussed at length.

The status of different subset-class cardinalities in ATMEMB will not be examined separately. This topic was being investigated already in connection with TMEMB. The observations provided in section 3.7.1.2 are relevant also with ATMEMB.

[46] In order to examine this modification, some comparisons groups were first compared with ATMEMB, then with a measure otherwise identical to it, but with the "reduced" divisor. Comparison groups #n/#n produced identical value groups. Value groups #n/#m with large differences between n and m, in contrast, were radically different. For example, the average of the ATMEMB value group #3/#9 was more than four times smaller than the corresponding figure for the reduced-divisor version.

3.7.2.2 ATMEMB Values of Z-related and Inversionally Related Pairs

As a total measure ATMEMB can discriminate between Z-related classes and, under T_n-classification, also between inversionally related classes. This gives rise to a number of interesting questions: Do all Z-pairs in a single comparison group share a uniform ATMEMB value?; Are there cases where a Z-related SC is more similar to some other class than to its Z-counterpart?; Under T_n-classification, are there cases where an inversionally non-symmetric SC is more similar to some other class than to its inversional counterpart?; etc.

Under T_n/I-classification, the highest ATMEMB value to be found among Z-related pairs, 0.91, belongs to the pair {8-Z15,8-Z29}. The lowest value, 0.65, belongs to two hexad class pairs, {6-Z4,6-Z37} and {6-Z26,6-Z48}. The average is 0.74. The Z-related hexad class pair with the highest value, 0.79, is {6-Z11,6-Z40}. Among the three Z-related pentad class pairs the values vary from 0.73 to 0.71, and among their 7-pc complement classes from 0.83 to 0.82. The sole tetrad class pair {4-Z15,4-Z29} gets the value 0.73. Comparing these values to those in the ATMEMB value group information table 3.14 reveals that they are usually above averages, but not on par with the highest values.

Every Z-related class produces its highest ATMEMB value with a class other than its Z-counterpart. This is extremely interesting, of course, as it suggests that identical dyad class contents do not in any way guarantee out of the ordinary similarity between subset-class contents of larger cardinalities. For example, the ATMEMB value between the Z-related pair {6-Z10,6-Z39} is 0.71. The individual ATMEMB value group 6-Z10/#2-#12 contains no less than 51 values exceeding 0.71. Among these: ATMEMB(6-Z10,7-Z37) = 0.86; ATMEMB(6-Z10,6-Z11) = 0.84; ATMEMB(6-Z10,5-Z18) = 0.77; etc. It is also noteworthy that numbers of values exceeding the Z-value may be different for the two counterparts. Let the other class in the pair {6-Z10,6-Z39} be our example. The individual ATMEMB value group 6-Z39/#2-#12 contains not 51, but only 36 values exceeding the Z-value 0.71. Moreover, the SCs producing high values with 6-Z10 may be considerably more distant from 6-Z39. For example, ATMEMB(6-Z39,7-Z37) = 0.65 (with 6-Z10 the 7-Z37 value was 0.86). ATMEMB(6-Z39,5-Z18) = 0.59 (with 6-Z10 0.77). Etc.

Generally, given in turn each Z-related SC pair $\{Z_1,Z_2\}$ with the two individual ATMEMB value groups Z_1/#2-#12 and Z_2/#2-#12, the average number of values exceeding the ATMEMB(Z_1,Z_2) value in each of the two value groups is 20. The smallest individual number of values above the Z-value is only two, belonging to classes in the pair {8-Z15,8-Z29}. According to ATMEMB - and, as we will see, also according to the other total measures - the Z-related classes turn out to be quite something else than the Siamese twins their dyad class contents deem them.

Under T_n-classification, pairs of inversionally related classes belonging to a single comparison group do not produce a uniform ATMEMB value. The only exception is the group #3/#3. (Table 3.15). All other total measures produce results repeating this extremely important observation.

The average ATMEMB value of all I-pairs is 0.76. Out of these 128 pairs, only 29 contain classes which are each other's closest ATMEMB counterparts. The pair {6-30A,6-30B}, for example, has the ATMEMB value 0.53. The individual ATMEMB value group 6-30A/#2-#12 contains an astounding 112 values exceeding 0.53. For 6-30B the corresponding number is the same.

Generally, given in turn each inversionally non-symmetric SC X, its inversionally related class I(X) and its individual ATMEMB value group X/#2-#12, the average number of values exceeding the ATMEMB(X,I(X)) value is 11.

TABLE 3.15: ATMEMB values indicating the highest, lowest and average degrees of similarity among pairs of inversionally related SCs. The six columns list (1) the comparison groups #n/#n, (2) the most similar I-pairs in the comparison groups, (3) the values belonging to the pairs in column 2, (4) the most dissimilar I-pairs in the comparison groups, (5) the values belonging to the pairs in column 4, (6) the average values of all I-pairs in the comparison groups.[47]

c.group	most simil.	value	most dissimil.	value	average
#9/#9:	{9-4A,9-4B}	0.95	{9-5A,9-5B}	0.93	0.94
#8/#8:	{8-19A,8-19B}	0.91	{8-Z15A,8-Z15B}	0.79	0.85
#7/#7:	{7-31A,7-31B}	0.94	{7-21A,7-21B}	0.77	0.81
#6/#6:	{6-14A,6-14B}	0.86	{6-30A,6-30B}	0.53	0.72
#5/#5:	{5-21A,5-21B}	0.88	{5-31A,5-31B}	0.58	0.71
#4/#4:	{4-22A,4-22B}	0.73	{4-Z15A,4-Z15B}	0.55	0.64
#3/#3:	All inversionally related 3-pc classes share the value 0.75				

Given an inversionally symmetric SC S and inversionally non-symmetric SCs X and Y, it is always so that ATMEMB(S,X) = ATMEMB(S,I(X)) and that ATMEMB(X,Y) = ATMEMB(I(X),I(Y)). ATMEMB(X,Y) may or may not be ATMEMB(X,I(Y)).[48]

3.7.2.3 ATMEMB: Conclusions

ATMEMB fulfils many important evaluation criteria and is based on a principle that

[47] If there are several I-pairs sharing a minimum or maximum value, only one pair is given as a representative of the pairs. This remark is valid also for tables 3.17 and 3.19 below.

[48] We do not have proofs for these formalisations. They, and the ones like them in connection with the other measures fulfilling C5, were obtained through exhaustive computer searches.

in certain types of case seems to produce more satisfactory results than measures based on one-to-one correspondence. (Section 3.6.10.2). In other types of case, however, the results it produces seem to be counterintuitive. In our opinion, the divisor term in the ATMEMB equation is flawed, resulting in values suggesting suspiciously high degrees of dissimilarity for SCs of clearly different cardinalities. The general reliability and usefulness of the measure is difficult to determine.

3.7.3 Lewin: REL

Presented in Lewin (1979-80b). A total measure pairing subset-class instances by one-to-many correspondence.

COMPARISON PROCEDURE:

To calculate REL(X,Y), the n-class vectors of X and Y, n ranging from 2 to the lesser of #X, #Y, are first obtained.[49] Two intermediate values are then calculated to reach the final value. (1) In each pair of corresponding vectors, corresponding components are multiplied. The square root is taken of each product. All square roots from all vector pair comparisons are added togeher. (2) All components in the vectors of X are added together and multiplied by the sum of all components in the vectors of Y. The square root of the product is taken. Finally, the intermediate value (1) is divided by the intermediate value (2).

EQUATION:[50]

Given SCs X and Y, the family TEST of all SCs of cardinalities 2 to the lesser of #X,#Y, the value of the function EMB(A,X), being the number of instances of SC A in X, and the value of the function TOTAL(X), being the number of all TEST class instances in X,

$$\mathrm{REL}(X,Y) = \frac{V}{W}$$

where

$$V = \sum_{A \in \mathrm{TEST}} \sqrt{\mathrm{EMB}(A,X) * \mathrm{EMB}(A,Y)} \quad \text{and} \quad W = \sqrt{\mathrm{TOTAL}(X) * \mathrm{TOTAL}(Y)}$$

[49] Lewin discusses REL at a very general level. The use of n-class vectors, for example, is not specifically required. This formulation, then, is our own.

[50] After Lewin (1979-80b:500) and Morris (1987:107). Morris's equation is incorrect as it suggests that EMB(A,X) and EMB(A,Y) in V are to be added together.

EXAMPLE 3.43: REL(5-7,5-15).[51] The prime forms are {0,1,2,6,7} and {0,1,2,6,8}, respectively. As the upper limit for TEST is 5, the nCVs to be compared are 2CVs, 3CVs, 4CVs and 5Cvs. TOTAL(5-7) is the sum of all components in 2CV(5-7), 3CV(5-7), 4CV(5-7) and 5CV(5-7). This sum, 26, is the same for 5-15. The value of the divisor W is √(26*26) = 26. To solve the value of the term V, we multiply the corresponding components, take the square root of each product and add all square roots together.

```
2CV(5-7)  =[ 3   1   0   1   3   2]
2CV(5-15)=[ 2   2   0   2   2   2]
  Σ√xiyi  =  √6+√2  +   √2+√6+√4

≈ 2.4 + 1.4 + 1.4.+ 2.4 + 2 ≈ 9.73.
```

```
3CV(5-7)  =[1  0  0  2  5  0  0  1  1  0  0  0]
3CV(5-15)=[1  0  0  2  2  0  0  4  1  0  0  0]
 Σ√xiyi   =√1     +    √4+√10    +    √4+√1

≈ 1 + 2 + 3.2 + 2 + 1 ≈ 9.16
```

```
4CV(5-7)  =[0 0 0 0 1 1 0 1 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0]
4CV(5-15)=[0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 1 0 0 0 0]
  Σ√xiyi  =         √2              +          √2                      ≈2.83
```

```
5CV(5-7)  =[0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
5CV(5-15)=[0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 Σ√xiyi   =                              0
```

V/W = (9.73 + 9.16 + 2.83 + 0)/26 = 21.72/26 ≈ 0.84.

EVALUATION CRITERIA FULFILLED:

C1, C2, C3.1, C3.2, C3.4, C4, C5.

THE REL VALUE GROUP #2-#12/#2-#12:

Non-integer values rounded to two decimal places. Value indicating highest degree of similarity: 1. Value indicating highest degree of dissimilarity: 0. Average: 0.57. Number of distinct values: 91. REL(X,Y) may or may not be REL(X_C,Y_C).

---

51 Same SCs were compared with TMEMB in Ex. 3.40 and ATMEMB in Ex. 3.42.

TABLE 3.16: The REL value groups #n/#m, $3 \leq n,m \leq 9$. Higher values indicate higher degrees of similarity. Each table cell contains, clockwise from the top left: the lowest and highest values, the number of distinct values, the average.

	#3	#4	#5	#6	#7	#8	#9
#3	0.0 0.6 0.28 6						
#4	0.0 0.99 0.4 18	0.13 0.85 0.45 38					
#5	0.16 0.88 0.45 29	0.12 0.92 0.52 67	0.19 0.86 0.58 54				
#6	0.15 0.85 0.45 42	0.1 0.88 0.51 61	0.17 0.98 0.6 70	0.14 0.88 0.63 61			
#7	0.2 0.67 0.45 32	0.23 0.7 0.5 42	0.29 0.8 0.59 48	0.25 0.93 0.65 64	0.45 0.91 0.71 41		
#8	0.19 0.57 0.43 25	0.23 0.66 0.47 37	0.26 0.74 0.56 41	0.22 0.88 0.63 54	0.45 0.98 0.72 49	0.5 0.91 0.76 38	
#9	0.24 0.52 0.42 18	0.21 0.58 0.45 30	0.31 0.62 0.52 26	0.27 0.73 0.59 36	0.49 0.86 0.71 35	0.45 0.97 0.79 39	0.7 0.94 0.86 17

3.7.3.1 Analysis

In (1977) and (1979-80b) Lewin suggests a probabilistic approach to various aspects of pcset theory. REL is a part of this discussion, being given as an alternative for Rahn's ATMEMB. For each member A of TEST, we define $p^X(A)$ = EMB(A,X)/TOTAL(X). Then p^X is a probability function on TEST. $p^X(A)$ is our expectation that, if a subset-class of X belonging to TEST is extracted, it will be specifically A. Given another SC Y, we can construct the analogous probability function p^Y on TEST. By correlating the individual results we can then evaluate the "relatedness" of X and Y with respect to the TEST family (after Lewin 1979-80b:499-501).

Lewin outlines three versions of REL. The one described above, being our norm, is the most extensive version comparing entire subset-class contents. Another version resembles $MEMB_n$ in that it compares only one subset-class cardinality at a time, but allows the cardinality to be selected freely. In the most limited version only 2CVs are compared. We will use this version, to be called REL_2 after Isaacson (1992:59), during comparisons with the measures based on one-to-one correspondence. (Section 3.7.3.4).

REL value 0, indicating maximal dissimilarity, can take place only if X and Y do not share any subset-classes. Value 1, indicating maximal similarity, is produced only if X = Y. REL resembles $MEMB_n$ and its derivatives in the respect that corresponding components can contribute to similarity only if both are nonzero. In multiplying the components the former has an advantage over the latter, however.

$MEMB_n$ needs a separate condition deeming component sum (x_i+y_i) valid only if $x_i > 0$ and $y_i > 0$. Lewin calls this condition "an arithmetic awkwardness." (1979-80b:501). He also notes that multiplying and taking the square root of the product produces the *geometric mean* of the components. Given components x_i and y_i, x_i is to the mean $\sqrt{(x_i y_i)}$ as the mean is to y_i (Ibid., 500).

Comparing the REL Table 3.16 to the ATMEMB Table 3.14, we see that the two one-to-many correspondence measures with normalised values produce rather similar results for comparison groups #n/#m where n = m or where the difference between n and m is small. For example, the lowest, highest and average values in the REL value group #3/#3 are 0, 0.6, and 0.28, respectively. The corresponding figures in the ATMEMB value group #3/#3 are 0, 0.62 and 0.29. Comparison groups #n/#m with large differences between n and m, in contrast, produce highly different REL and ATMEMB value groups. This is due to the peculiar feature in the divisor of the ATMEMB formula. (Section 3.7.2.1).

In section 3.7.1.2 above we examined the status of the different subset-class cardinalities in TMEMB by comparing the TMEMB and $MEMB_8$ value groups of the comparison group #9/#9. The $MEMB_8$ values, being incorporated in their TMEMB counterparts, were seen to contribute very little to the latter values. Since REL also processes all subset-classes as one entity, the TEST family of SCs, it shares the property of giving a very small representation for some subset-class cardinalities. (See related discussion in section 2.5).

3.7.3.2 REL Values of Z-related and Inversionally Related Pairs

Like TMEMB and ATMEMB, REL too can discriminate between Z-related classes and, under T_n-classification, between inversionally related classes. Under T_n/I-classification the Z-related pairs with the highest and lowest values are the same as they were with ATMEMB: REL(8-Z15,8-Z29) = 0.9, REL(6-Z4,6-Z37) = REL(6-Z26,6-Z48) = 0.64. The average is 0.73. Out of the Z-related hexad class pairs, REL too deems the pair {6-Z11,6-Z40} most similar. The value is 0.78. The Z-related septad classes get values between 0.83 and 0.81, their 5-pc complements between 0.73 and 0.7. REL(4-Z15,4-Z29) = 0.73.

Given in turn each Z-related SC pair $\{Z_1,Z_2\}$ with the two individual REL value groups Z_1/#2-#12 and Z_2/#2-#12, the average number of values exceeding the $REL(Z_1,Z_2)$ value in each of the two value groups is 30. The highest individual number of values above the Z-value belongs to 6-Z10. The REL value group 6-Z10/#2-#12 contains 66 values exceeding the REL(6-Z10,6-Z39) value. As with

ATMEMB, the numbers are not necessarily identical for the Z-counterparts. In the individual REL value group 6-Z39/#2-#12, 54 values exceed the REL(6-Z39,6-Z10) value. The SCs with the lowest number of values above the Z-value are again 8-Z15 and 8-Z29. They are the fourth closest classes with respect to one another. The results correspond quite closely with those produced by ATMEMB.

Under T_n-classification, pairs of inversionally related classes in a single comparison group produce varying REL values. Again, the group #3/#3 is an exception. Table 3.17.

TABLE 3.17: REL values indicating the highest, lowest and average degrees of similarity among pairs of inversionally related SCs. The six columns list (1) the comparison groups #n/#n, (2) the most similar I-pairs in the comparison groups, (3) the values belonging to the pairs in column 2, (4) the most dissimilar I-pairs in the comparison groups, (5) the values belonging to the pairs in column 4, (6) the average values of all I-pairs in the comparison groups.

c.group	most simil.	value	most dissimil.	value	average
#9/#9:	{9-4A,9-4B}	0.94	{9-8A,9-8B}	0.92	0.93
#8/#8:	{8-19A,8-19B}	0.90	{8-Z15A,8-Z15B}	0.79	0.84
#7/#7:	{7-31A,7-31B}	0.91	{7-21A,7-21B}	0.76	0.81
#6/#6:	{6-14A,6-14B}	0.86	{6-30A,6-30B}	0.53	0.71
#5/#5:	{5-21A,5-21B}	0.86	{5-31A,5-31B}	0.58	0.70
#4/#4:	{4-22A,4-22B}	0.73	{4-Z15A,4-Z15B}	0.55	0.64
#3/#3:	All inversionally related 3-pc classes share the value 0.75.				

The pairs in Table 3.17 are often the same as the corresponding ones in Table 3.15. Also the values correspond closely. The average I-pair value is 0.75. Out of these 128 pairs, only 11 are such that the classes are each other's closest REL counterparts. Among these are {6-14A,6-14B} and six 9-pc pairs. Generally, given in turn each inversionally non-symmetric SC X, its inversionally related class I(X) and its individual REL value group X/#2-#12, the average number of values exceeding the REL(X,I(X)) value is 18. The SCs with the highest number of values above the I-value form the pairs {4-Z15A,4-Z15B} and {4-Z29A,4-Z29B}. For each of these four SCs there are 131 classes closer than the I-counterpart. This figure is well over a third of all 352 transpositional SCs.

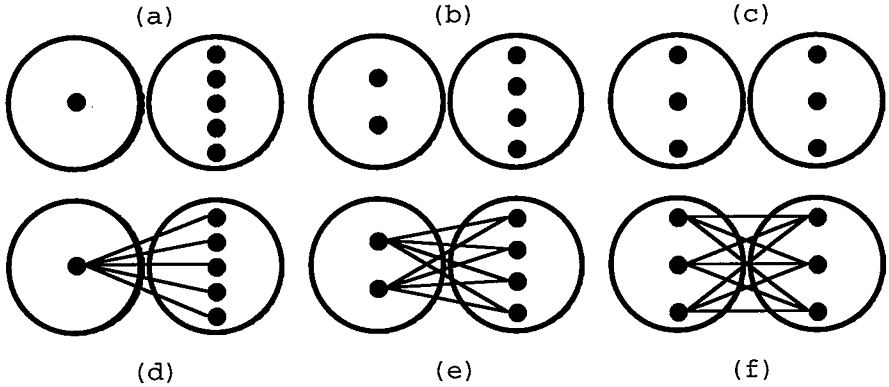
Given an inversionally symmetric SC S and inversionally non-symmetric SCs X and Y, it is always so that REL(S,X) = REL(S,I(X)) and that REL(X,Y) = REL(I(X),I(Y)). REL(X,Y) may or may not be REL(X,I(Y)).

3.7.3.3 REL and $MEMB_n$ as One-To-Many Correspondence Measures

As noted in section 3.6.10.2, multiplying corresponding vector components corresponds to determining the total number of (one-to-many) instance pairings between two SCs. The next step in REL, i.e., taking the square roots of the products to obtain the geometric means of the components, is necessary in order to keep peaked distributions from being rewarded.[52] Dividing the sum of the means, in turn, scales the values to make them lie between 0 and 1.

Let us examine which one of the two measures based on one-to-many correspondence, REL or $MEMB_n$, seems to produce more satisfying results. In Diagram 3.2, each circle represents a SC, the dots instances of some subset-class S. SC pair (a) is the same as pair (d), pair (b) same as pair (e), etc.

DIAGRAM 3.2:



$MEMB_n$ would produce a uniform degree of similarity for pairs (a)-(c). In each case, the sum of instances is six. REL, in turn, would suggest increasing similarity. The number of pairings, depicted as lines between the dots in pairs (d)-(e), increases from five to eight and then to nine. The corresponding geometric means are $\sqrt{5} \approx 2.24$, $\sqrt{8} \approx 2.83$ and $\sqrt{9} = 3$.

Of the two alternatives, i.e., $MEMB_n$ recording how many instances participate in the pairings and REL how many pairings there are, we prefer the latter. The pairings, we believe, are the true indicators of similarity. In $MEMB_n$, constant numbers of instances may create varying numbers of pairings, suggesting a concealed gradation of similarity.[53] In (c), six instances produce nine pairings. In (b), eight, and in (a), five.

52 Suppose we compare 2CV(3-3) = [1 0 1 1 0 0] to itself only by multiplying corresponding components. We get (1*1) + (1*1) + (1*1) = 3. Comparing 2CV(3-12) = [0 0 0 3 0 0] to itself would produce (3*3)=9, suggesting that 3-12 is considerably more similar to itself than 3-3 is to itself. Taking the geometric means equalizes the results, $\sqrt{1}+\sqrt{1}+\sqrt{1}=3$ and $\sqrt{9}=3$.

53 See a related observation concerning Forte's R_1 and R_2 relations in section 3.5.2.1.

3.7.3.4 REL and the One-To-One Correspondence Measures

Above examples 3.20 and 3.37 gave four SC pairs, each involving the triad class 3-5 and one of the tetrad classes 4-9, 4-6, 4-5 and 4-Z15. We suggested that this tetrad class order means increasing dissimilarity for the pairs. We criticized SIM, k, ASIM and ak for being unable to identify the gradation. $\%REL_2$ and $MEMB_n$, by contrast, produced values suggesting decreasing similarity. REL belongs to the latter category. REL(3-5,4-9) = 0.99; REL(3-5,4-6) = 0.83; REL(3-5,4-5) = 0.7; REL(3-5,4-Z15) = 0.63.

For further comparisons between REL and $MEMB_n$, as well as between REL and the one-to-one correspondence measures, let us return to the set-class pairs we already compared with $MEMB_2$, SIM and $\%REL_2$ in section 3.6.10.2. They are {3-3,3-8}, {3-3,3-12}, {4-1,4-28} and {4-3,4-9}. The prime forms, 2CVs and 2C%Vs are given in the examples 3.44 and 3.45. Instead of the total version of REL, we will use the smaller one, REL_2. Each measure, then, compares only 2CVs.

EXAMPLE 3.44: Prime forms 2CVs, and 2C%Vs of three SCs

```
3-3:  {0,1,4}  [1 0 1 1 0 0] [33 0 33  33 0  0]
3-8:  {0,2,6}  [0 1 0 1 0 1] [0 33  0  33 0 33]
3-12: {0,4,8}  [0 0 0 3 0 0] [0  0  0 100 0  0]
```

SIM and $\%REL_2$ suggested a uniform degree of similarity to the triad class pairs, whereas $MEMB_2$ deemed the pair {3-3,3-12} closer. We agreed with the latter result. Also REL_2 produces values indicating varying degrees of similarity. REL_2(3-3,3-8) = 0.33, REL_2(3-3,3-12) = 0.58. The former pair produces only one pairing between corresponding dyad class instances, the latter three.

EXAMPLE 3.45: Prime forms, 2CVs and 2C%Vs of four SCs

```
4-1:  {0,1,2,3}, [3 2 1 0 0 0], [50 33 17 0 0  0]
4-28: {0,3,6,9}, [0 0 4 0 0 2], [ 0  0 67 0 0 33]

4-3:  {0,1,3,4}, [2 1 2 1 0 0], [33 17 33 17  0  0]
4-9:  {0,1,6,7}, [2 0 0 0 2 2], [33  0  0  0 33 33]
```

According to SIM and $\%REL_2$, the pair {4-1,4-28} was more dissimilar than the pair {4-3,4-9}. $MEMB_2$ deemed the first pair closer than the second. This time we agreed with the former result. REL_2 offers a compromise: REL_2(4-1,4-28) = REL_2(4-3,4-9) = 0.33. In both cases the number of pairings between corresponding dyad class instances is four, 1*4 in the former case and 2*2 in the latter. Here, in our opinion, the one-to-one correspondence measures reflect intuitive similarity better than is done by REL_2. The equal representation of 2-1 in both classes is the very aspect producing the sense of closer similarity to the pair {4-3,4-9}.

We repeat that both measure categories seem to have their advantages and disadvantages. In some cases one of the approaches seems to have the upper hand, only to lose it in others. The analysis in section 3.6.10.2 and here suggests that when compared to the one-to-many approach, one-to-one correspondence offers a sort of middle course. It may leave some aspects unnoticed in some cases, but does not produce clearly counterintuitive results, either. This might result in higher overall reliability.

3.7.3.5 REL: Conclusions

On the basis of the analysis above, and also on the basis of the many important criteria it fulfils, we conclude that REL is a well-conceived and good measure of SC similarity. In our opinion the values it produces correlate well with intuitive sense of similarity: small-cardinality comparison groups produce wide ranges of values with averages suggesting high degree of dissimilarity; large-cardinality comparison groups produce relatively narrow ranges of values with averages suggesting high degree of similarity (Table 3.16); pairs with exceptional intuitive closeness get extremely high values: REL(6-35,5-33) = 0.98; etc. REL shares the advantages of ATMEMB, partially avoiding some of its disadvantages, while avoiding others completely.

3.7.4 Castrén: T%REL

Total Percentage Relation. An expanded version of the $\%REL_n$ measure. Presented in an unpublished manuscript. A total measure comparing proportionate subset-class contents. A one-to-one correspondence measure.

COMPARISON PROCEDURE:
Given SCs X and Y, the classes are compared with $\%REL_2$, $\%REL_3$,... $\%REL_m$. When X and Y are of different cardinalities, m is the lesser of #X, #Y. Otherwise, m is #X-1. The final result is the average of the $\%REL_n$ values, rounded to the nearest integer.

EQUATION:

Given SCs X and Y and the variable m = MIN(#X,#Y), if #X ≠ #Y, otherwise m = #X-1,

$$T\%REL(X,Y) = \frac{\sum_{n=2}^{m} \%REL_n(X,Y)}{(m-1)}$$

EXAMPLE 3.46: T%REL(5-2,5-3). The prime forms are {0,1,2,3,5} and {0,1,2,4,5}, respectively. T%REL(5-2,5-3) = (%REL$_2$(5-2,5-3) + %REL$_3$(5-2,5-3) + %REL$_4$(5-2,5-3))/3.

```
     2C%V(5-2)  =  [30 30 20 10 10 0]
     2C%V(5-3)  =  [30 20 20 20 10 0]

%REL2(5-2,5-3)  =  |30-30|+|30-20|+|20-20|+|10-20|+|10-10|+|0-0|
                                            2
                =  0+10+0+10+0+0  =  10
                         2

     3C%V(5-2)  =  [20 30 10 10 0 10 20 0 0 0 0 0]
     3C%V(5-3)  =  [10 20 30 20 0 10 10 0 0 0 0 0]

%REL3(5-2,5-3)  =  |20-10|+|30-20|+|10-30|+|10-20|+|0-0|+|10-10|+|20-10|
                                            2
                =  10+10+20+10+0+0+10  =  30
                           2

4C%V(5-2)=[20 20  0 20 0 0  0 0 0 20 20 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
4C%V(5-3)=[ 0 20 20 20 0 0 20 0 0  0 20 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]

%REL4(5-2,5-3)  =  |20-0|+|20-20|+|0-20|+|20-20|+|0-20|+|20-0|+|20-20|
                                            2
                =  20+20+20+20  =  40
                         2
```

T%REL(5-2,5-3) = (10+30+40)/3 ≈ 27.

EVALUATION CRITERIA FULFILLED:

C1, C2, C3.1, C3.2, C3.3, C3.4, C4, C5.

THE T%REL VALUE GROUP #2-#12/#2-#12:

All values are integers. Value indicating highest degree of similarity: 0. Value indicating highest degree of dissimilarity: 100. Average: 63. Number of distinct values: 79. T%REL(X,Y) may or may not be T%REL(X_C,Y_C).

TABLE 3.18: T%REL value groups #n/#m, 3 ≤ n,m ≤ 9. Lower values indicate higher degrees of simi - larity. Each table cell contains, clockwise from the top left: the lowest and highest values, the number of distinct values, the average. C = 100.

	#3	#4	#5	#6	#7	#8	#9
#3	33 C 68.35 3						
#4	0 C 74.97 10	21 92 62.53 12					
#5	27 95 74.37 19	28 93 69.7 52	27 83 56.28 18				
#6	27 93 74.36 31	28 93 68.66 49	0 88 63.56 54	25 88 54.88 52			
#7	47 93 74.21 30	45 91 67.88 43	39 85 61.97 41	26 88 59.79 53	25 70 48.35 41		
#8	59 93 74.2 23	48 90 67.48 40	45 85 61.36 35	26 88 58.72 46	0 75 53.41 47	23 68 44.87 42	
#9	64 91 74.25 20	51 90 67.07 34	51 83 60.78 28	43 87 58.01 35	32 66 51.69 30	14 71 47.83 39	24 54 36.36 20

3.7.4.1 Analysis

When the compared SCs X and Y are of the same cardinality m, their m-class %-vectors are not among the compared nC%Vs. The reason is that unless X = Y, $\%REL_m(X,Y)$ is automatically 100 and does not meaningfully participate in determining the degree of similarity. As T%REL(X,Y) is the average of the individual $\%REL_n(X,Y)$ values, the m-class value would only serve to increase the final outcome.

When comparing SCs of large cardinalities with total measures processing all subset-classes as one entity, the largest subset-class cardinalities were seen to have a weak impact in the values. (Sections 2.5, 3.7.1.2 and 3.7.3.1). As T%REL compares each subset-class cardinality separately, the situation is entirely different. Let us again examine the comparison group #9/#9.

Among the 66 values in the $\%REL_8$ value group #9/#9, there are 13 instances of the maximum value 100. The lowest value is 56, the average 83. When the T%REL value group #9/#9 is obtained, the octad class distributions, suggesting strong dis - similarity, contribute exactly as much to the final outcome as dyad class, triad class, and all the other distributions do. As a result, T%REL leads to an observation exactly contrary to that produced by $\%REL_2$. Some values are too *high* to be intuitively acceptable. See Table 3.18. For example, the lowest value in the value group #9/#9 is 24, the average ca. 36. This suggests a rather high degree of dissimilarity for nonad class pairs in general. The pentad classes are deemed still more distant from each other. The value group #5/#5 minimum is 27 and the average circa 56. The latter en -

try is twice as high as its counterpart for %REL_2. (Table 3.2). It is to be noted, however, that highly different cardinalities do not produce counterintuitively low values. The insensitivity to large cardinality differences we observed in %REL_2 is avoided here.

We believe that processing different subset-class cardinalities separately is a correct approach. The high values are a result from not taking the non-common subset-class criterion C6 into account. We saw above, for example, that disjoint octad class contents may bear some, possibly considerable, similarity. (Section 3.7.1.2). Relation to criterion C6 is exactly the difference between T%REL and RECREL. The former measures the extent to which the proportioned subset-class contents of two SCs overlap, and stops there. For the latter, this is merely the starting point. Degrees of similarity are determined also for the unilaterally embedded subset-classes of cardinalities 3 and higher. Consequently, RECREL values are lower than their T%REL counterparts.

3.7.4.2 T%REL Values of Z-related and Inversionally Related Pairs

As a total measure T%REL, like TMEMB, ATMEMB and REL, can discriminate between Z-related classes and, under T_n-classification, between inversionally related classes. Under T_n/I-classification the most similar and most dissimilar Z-related SCs are the same as with ATMEMB and REL. T%REL(8-Z15,8-Z29) = 23. T%REL(6-Z4,6-Z37) = T%REL(6-Z26,6-Z48) = 50. The average value is 39. Out of the Z-related hexad class pairs, the highest degree of similarity belongs to {6-Z44,6-Z19}, of value 30. Values of the Z-related septad classes vary between 31 and 36, those of their 5-pc complements between 30 and 40. T%REL(4-Z15,4-Z29) = 25.

Given in turn each Z-related SC pair {Z_1,Z_2} with the two individual T%REL value groups Z_1/#2-#12 and Z $_2$/#2-#12, the average number of values below the T%REL(Z_1,Z_2) value in each of the two value groups is 10. The class with most values below the Z-value is 6-Z39. The individual T%REL value group 6-Z39/#2-#12 contains 30 values below the T%REL(6-Z39,6-Z10) value. The value group 6-Z10/#2-#12, in turn, contains 27 such values. At the other end of the scale there are 14 SCs for which their very lowest T%REL values are with their Z-counterparts.

Under T_n-classification, T%REL produces results resembling those from the other total measures. With the exception of group #3/#3, inversionally related SCs within a single comparison group do not produce uniform values. Table 3.19.

The average I-pair value is 33. Generally, given in turn each inversionally non-symmetric SC X, its inversionally related class I(X) and its individual T%REL value group X/#2-#12, the average number of values below the T%REL(X,I(X)) value is four. This figure is lower than those produced by ATMEMB and REL, being 11 and 18, respectively. Furthermore, almost half of the inversionally related SC pairs, 62 out of 128, are such that the two I-counterparts are each other's closest T%REL counterparts. For ATMEMB and REL the corresponding figures were 29 and 11 out of 128, respectively. The SCs with the largest number of values below the I-counterpart value are 6-Z15A, 6-Z15B, 6-Z31A and 6-Z31B. The number is 35 in each case.

TABLE 3.19: T%REL values indicating the highest, lowest and average degrees of similarity among pairs of inversionally related SCs. The six columns list (1) the comparison groups #n/#n, (2) the most similar I-pairs in the comparison groups, (3) the values belonging to the pairs in column 2, (4) the most dissimilar I-pairs in the comparison groups, (5) the values belonging to the pairs in column 4, (6) the average values of all I-pairs in the comparison groups.

c.group	most simil.	value	most dissimil.	value	average
#9/#9:	{9-8A,9-8B}	20	{9-5A,9-5B}	21	21
#8/#8:	{8-19A,8-19B}	26	{8-Z15A,8-Z15B}	38	32
#7/#7:	{7-7A,7-7B}	23	{7-Z38A,7-Z38B}	38	33
#6/#6:	{6-14A,6-14B}	23	{6-30A,6-30B}	55	39
#5/#5:	{5-7A,5-7B}	20	{5-31A,5-31B}	47	35
#4/#4:	{4-22A,4-22B}	25	{4-Z15A,4-Z15B}	50	38
#3/#3:	All inversionally related 3-pc classes share the value 0				

Given an inversionally symmetric SC S and inversionally non-symmetric SCs X and Y, it is always so that T%REL(S,X) = T%REL(S,I(X)) and that T%REL(X,Y) = T%REL(I(X),I(Y)). T%REL(X,Y) may or may not be T%REL(X,I(Y)).

■ CHAPTER 4
RECREL

4.1 INTRODUCTION

In this chapter we will introduce the RECREL similarity measure, analyse its details and give some examples of comparisons between pairs of set-classes. Due to the complexity of the measure, we adopt an approach slightly different from the one we took with the other similarity measures. This time we will not lay out all its aspects and details straight away, but will proceed gradually from the general to the particular. We will go through the comparison procedure three times, each new "round" building on top of the aspects introduced in the previous one. The terms and concepts are described also in the glossary.

When we introduce the measure initially, in section 4.2, we will use the same sort of schematic representation as we did in section 2.4.5. Here our sole objective is to describe RECREL in broad outline and address the following questions: Why is $\%REL_n$, the internal similarity measure of RECREL, evaluated repeatedly throughout a single comparison?; What kind of further tasks would seem to arise from a $\%REL_n$ evaluation and how do these lead to still new evaluations?; How do the values in the tree-shaped net of $\%REL_n$ comparisons relate to one another?; How are the individual $\%REL_n$ values to be processed to get a final RECREL value?; etc.

In section 4.3, while going through the measure for the second time, we shall discuss and give names to a number of concepts already identified and used in the previous section. With the help of the new terminology, we analyse a real RECREL comparison between two pentad classes, step by step.

In the third and final pass through RECREL, in section 4.4, we will mostly discuss aspects concerning a proper way to derive a final RECREL value from the many $\%REL_n$ values. We will see that, while the comparison in section 4.3 produced correct results, the characteristics of some other comparisons require additional aspects to be taken into account. After we have identified and described these, the RECREL similarity measure will have been introduced in its entirety. The section that follows, 4.5, is devoted to a detailed example, a comparison between a septad class and a tetrad class.

The nature of section 4.6 is different from those of the sections preceeding it. We describe the whole comparison mechanism as concisely as possible, first as a ten-step algorithm and then as a recursive Common Lisp function. In the final section, 4.7, we examine the relation between RECREL and the similarity measure criteria given in section 2.3.

4.2 RECREL IN OUTLINE

Let us recall the schematic example we discussed in connection with criterion C6 in section 2.4.5, and extend it somewhat in order to see the outline of a RECREL comparison.

We have two septad classes, 7-X and 7-Y, and want to measure the degree of similarity between them. In doing this, our basic tool will be the %REL_n similarity measure. From the indexes of the nonzero components in the 6-class %-vectors of 7-X and 7-Y, we infer that the hexad classes contained in 7-X are 6-A, 6-B and 6-C, those in 7-Y 6-C, 6-D and 6-E. As the only mutually embedded hexad class is 6-C, the value for %REL_6(7-X,7-Y) would in all probability be high.

We have so far established the extent to which the proportioned hexad class contents of 7-X and 7-Y consist of similar elements. A notion we know nothing about as yet, however, is how the non-shared (unilaterally embedded) hexad classes in 7-X, 6-A and 6-B, are related to those in 7-Y, 6-D and 6-E. If we were to deem them highly similar, the sense of similarity between 7-X and 7-Y would be strengthened. If they turned out to be highly dissimilar, the high %REL_6(7-X,7-Y) value would be corroborated.

Assessing this degree of similarity requires some effort, as there are four SCs involved. We have to compare each unilaterally embedded 6-pc class in 7-X to every corresponding class in 7-Y. The number of pairs to be examined, then, is four: {6-A,6-D}, {6-A,6-E}, {6-B,6-D} and {6-B,6-E}. For the sake of brevity, let us concentrate on only one of them, {6-A,6-D}.

We compare the two hexad classes with %REL_5. The indexes of the nonzero components in 5C%V(6-A) and 5-C%V(6-D) show that the pentad classes contained in 6-A are 5-A, 5-B and 5-C, whereas those in 6-D are 5-C, 5-D and 5-E. Only 5-C is mutually embedded, suggesting a high value for %REL_5(6-A,6-D). 5-A and 5-B, the unilaterally embedded pentad classes in 6-A, are by definition different classes than 5-D and 5-E, their counterparts in 6-D. If a cross-correlated comparison of these SCs is made, each pair represents some degree of similarity, which, in turn, can be used to assess the degree of similarity between 6-A and 6-D more accurately. The four pairs are {5-A,5-D}, {5-A,5-E}, {5-B,5-D} and {5-B,5-E}.

We examine the first pair only, comparing it with %REL4. Let the 4-pc subset-classes of 5-A be 4-A, 4-B and 4-C and those of 5-D 4-C, 4-D and 4-E. The only mutually embedded class is 4-C, suggesting a high value for %REL4(5-A,5-D). As the unilaterally embedded tetrad classes, 4-A and 4-B in 5-A, 4-D and 4-E in 5-D, can participate in assessing the degree of similarity between 5-A and 5-D, they have to be paired for further cross-correlated comparisons. The four pairs are {4-A,4-D}, {4-A,4-E}, {4-B,4-D} and {4-B,4-E}.

The ever-expanding tree of comparisons has two more levels to go. As an example of the tetrad class pairs we compare 4-A and 4-D with %REL3, observing also their 3-pc subset-class contents. These are {3-A,3-B,3-C} and {3-C,3-D,3-E}, respectively. We combine the unilaterally embedded triad classes into pairs {3-A,3-D}, {3-A,3-E}, {3-B,3-D} and {3-B,3-E}, and, finally, compare the dyad class contents of the SCs in these pairs with %REL2.

Subset-classes of cardinality 2 are the smallest we examine. Consequently, the four values from %REL2(3-A,3-D), %REL2(3-A,3-E), %REL2(3-B,3-D) and %REL2(3-B,3-E) are accepted as they are, without further deliberations on possible similarities between unilaterally embedded dyad classes. Together, these values determine the degree of similarity between the triad classes embedded unilaterally in 4-A and 4-D. Next, we want the four values to be processed so that the result is a single value representing them. For the time being, let us take their average and call it *a*.[1] Besides *a*, 4-A and 4-D produced another value, the one from %REL3(4-A,4-D). Let us call this original value *o*.

We have now to decide what is the mutual relationship of these values. Shall we, for example, abandon *o* and determine *a* to be the final value? This is not possible, because *o* gives the original extent to which 3C%V(4-A) and 3C%V(4-D) differ, and *a* is tied to it. We expressly need *o* to indicate how large is the "dissimilar share" whose internal degree of similarity is *a*. Quite simply, *a* percent from a large share is different from *a* percent from a small one. The original value can be anything at all, of course, 5, 50 or 95.

We decide that we *update* the original value *o* by taking *a* percent from it: $(o * a)/100$. In the tree of comparisons, this is the final value for the pair {4-A,4-D}. It

[1] Later on, after some new concepts have been introduced, the way to determine a value representing a set of values will be slightly modified.

reflects two separate notions: first, how large are the shares that the unilaterally embedded classes have in the entire 3-pc subset-class contents of 4-A and 4-D, and second, what is the internal degree of similarity between those two groups of unilaterally embedded classes. If both *o* and *a* were 90, for example, the final value would be (90 * 90)/100 = 81. The very high degree of dissimilarity between the unilaterally embedded SCs would change the original value only slightly. If, on the other hand, *o* were to be 90 and *a* 20, the final {4-A,4-D} value would be 18. Here the non-shared 3-pc subset-classes, being similar to one another, decrease the final degree of similarity considerably.

In an identical manner, groups of triad class pairs are derived also from each of the three "sibling" pairs of {4-A,4-D}. They are {4-A,4-E}, {4-B,4-D} and {4-B,4-E}. Each group produces an average value updating its "parent" tetrad class pair value. The four updated tetrad class values, in turn, together give the degree of similarity between the unilaterally embedded tetrad class materials of 5-A and 5-D, their own "parent" pair. We take the average *a* of these four values and update the %REL4(5-A,5-D) value *o* with it. The new value takes into account both the shares of the unilaterally embedded 4-pc subset-classes in the tetrad class contents of 5-A and 5-D, and the degree of similarity between those two disjoint groups of SCs. Each of the "sibling" pairs of {5-A,5-D}, i.e., {5-A,5-E}, {5-B,5-D} and {5-B,5-E}, generates its own tree of $\%REL_n$ comparisons and from this, in reversed order, a tree of value updatings where the value of each SC pair is updated with the values of its "offspring" pairs. When all four pentad class pairs have received their updated values, the average *a* of these values updates the original "parent" %REL5(6-A,6-D) value *o*. The other hexad class pairs, {6-A,6-E}, {6-B,6-D} and {6-B,6-E}, receive their own updated values from corresponding trees of comparisons and value updatings. When the average *a* of the four hexad class pair values updates the original %REL6(7-X,7-Y) value *o*, the succession of comparisons and value updatings is complete.

In the example above, let us assume that each SC in each compared pair contains two unilaterally embedded subset-classes. When paired cross-correlatedly, the subset-classes of each pair would then produce four new pairs to be compared. Altogether the number of %REL6 evaluations is one, the number of %REL5 evaluations 4, the number of %REL4 evaluations 16, the number of %REL3 evaluations 64 and the number of %REL2 evaluations 256. The total number of $\%REL_n$ evaluations in our example, then, would be 341. Obviously, the "REC" in RECREL is for *recursive*, reflecting the tendency of the measure to produce many $\%REL_n$ evaluations from one initial evaluation.[2]

Despite the large number of $\%REL_n$ evaluations, the steps taken so far do not constitute the entire RECREL comparison between 7-X and 7-Y, only the major part of it. We have not, for example, compared the pentad class contents of the two septad classes yet, only those of the unilaterally embedded hexad classes. These comparisons are entirely separate and may produce altogether different results. In order to make the RECREL(7-X,7-Y) comparison complete, we have to generate similar trees of comparisons and value updatings from all mutually embeddable subset-class cardinalities larger than one. The "top-level" comparisons of these trees, from which the lower-level evaluations are derived, are $\%REL_5$(7-X,7-Y), $\%REL_4$(7-X,7-Y), $\%REL_3$(7-X,7-Y) and $\%REL_2$(7-X,7-Y). With the exception of the last one, each would produce further comparisons at a lower level or levels. The final RECREL value is the average of the individual values.

4.3 TERMINOLOGY AND AN EXAMPLE

4.3.1 Terminology

4.3.1.1 Difference Vectors

The RECREL outline in section 4.2 was slightly simplified in certain respects. For example the mutually embedded subset-classes, such as 6-C in both 7-X and 7-Y, were always excluded from further comparisons. In an actual comparison, however, we would have had to observe not only whether a given subset-class is mutually embedded, but also how strongly it is represented in the two %-vectors. For example the component belonging to 6-C in 6C%V(7-X) might have been larger than its counterpart in 6C%V(7-Y). That is, 7-X could have contained a unilaterally embedded *share* of 6-C. When the differences between corresponding 6C%V components were found out during the $\%REL_6$(7-X,7-Y) comparison, this excess share of 6-C would have contributed to dissimilarity in exactly the same manner as the nonzero components with zero counterparts, belonging to 6-A and 6-B in 6C%V(7-X) and 6-D and 6-E in 6C%V(7-Y).

Because mutually embedded subset-classes may have different representations in their respective superset-classes, we have to refine the way with which the unilaterally embedded subset-class materials are identified for further comparisons.

---

[2] See glossary entries "recursive" and "recursive routine".

The proper way resembles a %REL$_n$ comparison. Given the n-class %-vectors of SCs X and Y, we take the differences between pairs of corresponding components $\{c_X, c_Y\}$ in both ways, $(c_X - c_Y)$ and $(c_Y - c_X)$. Each non-negative, nonzero difference $(c_X - c_Y)$ indicates that some subset-class has a larger representation in X than in Y. The non-negative, nonzero differences $(c_Y - c_X)$ indicate larger representations in Y. We retain only the non-negative differences, replacing the negative ones with zeros. The two resulting *difference vectors* give the unilaterally embedded subset-class shares in X and Y.[3]

EXAMPLE 4.1: The 3-class %-vectors of SCs 4-2A and 4-2B (rows 1 and 2 from top, respectively), an index row and the difference vectors of 4-2A and 4-2B, respectively.

```
[25 25  0 25  0  0  0  0  0 25  0  0  0  0  0  0  0  0  0]
[25  0 25  0 25  0  0  0  0 25  0  0  0  0  0  0  0  0  0]

  1  2  2  3  3  4  4  5  5  6  7  7  8  8  9 10 11 11 12

[ 0 25  0 25  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
[ 0  0 25  0 25  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
```

4.3.1.2 Difference Groups

Given the comparison %REL$_n$(X,Y), the *difference groups* of X and Y consist of the subset-classes of cardinality n having unilaterally embedded shares in X and Y, respectively.[4] The names of the subset-classes are to be inferred from the indexes of nonzero components in two difference vectors. In Ex. 4.1, the difference group of 4-2A is {3-2A,3-3A}, that of 4-2B {3-2B,3-3B}. Two difference groups derived from a single comparison can never contain the same SCs.

4.3.1.3 Cross-Correlation Groups

Given the comparison %REL$_n$(X,Y) within RECREL so that $n > 2$, each subset-class in the difference group of X must be compared to every class in that of Y with %REL$_{n-1}$. If the two difference groups contain n and m SCs, respectively, the resulting *cross-correlation group* contains $(n * m)$ SC pairs. The cross-correlation group

[3] As the negative differences are not saved and as the starting points can be any n-class %-vectors, our difference vectors are not similar to the Interval-difference Vectors in Isaacson (1990) and (1992).

[4] The concept will be slightly expanded later on, as each subset-class in a difference group will be associated with its *weight*.

to be derived from Ex. 4.1 contains the pairs {{3-2A,3-2B},{3-2A,3-3B}, {3-3A,3-2B}, {3-3A,3-3B}}.[5]

4.3.1.4 Branches

Within the schematically represented RECREL comparison in section 4.2, we examined an entity consisting of three components: the top-level comparison $\%REL_6$(7-X,7-Y), the tree of comparisons derived from it, and the updatings of the comparison values. An entity like this will be called a *branch*. Being an independent part of a RECREL comparison, a branch is numbered after the n in the top-level $\%REL_n$. The branch in section 4.2, then, is six, after $\%REL_6$(7-X,7-Y). The next branch is five, containing $\%REL_5$(7-X,7-Y), the tree of comparisons derived from it and the value updatings. Branch four follows, etc., until branch two, consisting of only the comparison $\%REL_2$(7-X,7-Y), has received its value. The final RECREL value is the average of the branch values.

When two SCs X and Y of cardinality n are compared with RECREL, the branches are n-1, n-2,..2. Examination of branch n is not meaningful, since $\%REL_n(X,Y) = 100$ for every pair {X,Y} where $X \neq Y$. The cross-correlation group derived from $\%REL_n$(X,Y) would consist only of the pair {X,Y} itself, meaning that $\%REL_{n-1}$(X,Y) and the comparisons following it would exactly duplicate the branch n-1. Pairs of dyad classes are exceptions to this rule. They are compared with $\%REL_2$, since the lowest branch to be calculated, two, is the only one possible for them.

When SCs X and Y of cardinalities n and m, $n < m$, are compared with RECREL, the branches are n, n-1, n-2,..2. Thus, for example, the branches within a RECREL comparison between a 9-pc SC X and a 4-pc SC Y are four, three and two. Now, as the value for $\%REL_4$(X,Y) is not automatically 100, Y can meaningfully participate in the comparison.

4.3.1.5 Levels

Given the branch n, $n > 2$, within the RECREL comparison between SCs X and Y, we say that the first comparison $\%REL_n$(X,Y) takes place at *level* n. The $\%REL_{n-1}$ comparisons of the pairs in the cross-correlation group derived from X and Y take place

[5] Due to the difference group weights to be adopted later on, the cross-correlation group pairs will also be associated with some additional information.

in branch n, at level n-1. Accordingly, the group can be identified as a level-n-1 cross-correlation group. The lowest level is two, comprising the %REL2 comparisons of the pairs in the one, or more, level-two cross-correlation groups.

4.3.2 An Example: RECREL(5-2A,5-2B)

In our first example of a genuine RECREL comparison, the classes to be compared are 5-2A, {0,1,2,3,5}, and its inversionally related SC 5-2B, {0,2,3,4,5}. The branches are four, three and two.

EXAMPLE 4.2.1: The 4-class %-vectors of 5-2A (top row) and 5-2B (second row), a row of indexes and the difference vectors of 5-2A and 5-2B, respectively. %REL4(5-2A,5-2B) = 60.

[20 20 0 0 20 0 0 0 0 0 0 0 20 20 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
[20 0 20 0 0 20 0 0 0 0 0 0 20 0 20 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]

1 2 2 3 4 4 5 5 6 7 8 9 10 11 11 12 12 13 13 14 14 15 15 16 16 17 18 18 19 19 20 21 22 22 23 24 25 26 27 27 28 29 29

[0 20 0 0 20 0 0 0 0 0 0 0 0 20 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
[0 0 20 0 0 20 0 0 0 0 0 0 0 0 20 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]

The levels in branch four are four, three and two. The value of the top-level comparison %REL4(5-2A,5-2B) is 60. The two difference groups derived from the difference vectors in Ex.4.2 are as follows: {4-2A,4-4A,4-11A} for 5-2A, {4-2B,4-4B,4-11B} for 5-2B. The nine-pair level-three cross-correlation group from these is {{4-2A,4-2B}, {4-2A,4-4B}, {4-2A,4-11B}, {4-4A,4-2B}, {4-4A,4-4B}, {4-4A,4-11B}, {4-11A,4-2B}, {4-11A,4-4B}, {4-11A,4-11B}}. The pairs will be compared with %REL3.

Besides a %REL3 value, each cross-correlation group pair produces two difference vectors, two difference groups and, as a combination of the latter, a level-two cross-correlation group. The 3C%Vs and difference vectors of one of these pairs, {4-2A,4-2B}, is given in Ex. 4.1 above. The corresponding difference groups are given in section 4.3.1.2 and the resulting cross-correlation group in section 4.3.1.3.

The triad class pairs in the nine level-two cross-correlation groups are compared with %REL2. Once this is completed, we have made 86 %RELn comparisons at three different levels within branch four. The pairs participating in the comparisons, as well as the resulting values, are given as a diagram in Ex. 4.2.2. The one and only level-four comparison, between the pentad classes, is given first. The level-three cross-correlation group, containing tetrad class pairs, is given as a column, each pair being followed by the level-two cross-correlation group derived from it. The multiple parentheses and the indentation also serve to indicate which list cells belong together.[6]

EXAMPLE 4.2.2: RECREL(5-2A,5-2B), branch four. The 86 %REL_n comparisons at levels four, three and two.

```
((5-2A,5-2B,60.0)
  (((4-2A,4-2B,50.0)
    ((3-2A,3-2B,0.0),(3-2A,3-3B,33.3),(3-3A,3-2B,33.3),(3-3A,3-3B,0.0)))
   ((4-2A,4-4B,75.0)
    ((3-2A,3-3B,33.3),(3-2A,3-4B,66.7),(3-2A,3-7B,33.3),(3-3A,3-3B,0.0)
     (3-3A,3-4B,33.3),(3-3A,3-7B,66.7),(3-6,3-3B,66.7),(3-6,3-4B,66.7)
     (3-6,3-7B,66.7)))
   ((4-2A,4-11B,75.0)
    ((3-1,3-2B,33.3),(3-1,3-4B,66.7),(3-1,3-7A,66.7),(3-2A,3-2B,0.0)
     (3-2A,3-4B,66.7),(3-2A,3-7A,33.3),(3-3A,3-2B,33.3),(3-3A,3-4B,33.3)
     (3-3A,3-7A,66.7)))
   ((4-4A,4-2B,75.0)
    ((3-3A,3-2B,33.3),(3-3A,3-3B,0.0),(3-3A,3-6,66.7),(3-4A,3-2B,66.7)
     (3-4A,3-3B,33.3),(3-4A,3-6,66.7),(3-7A,3-2B,33.3),(3-7A,3-3B,66.7)
     (3-7A,3-6,66.7)))
   ((4-4A,4-4B,75.0)
    ((3-3A,3-3B,0.0),(3-3A,3-4B,33.3),(3-3A,3-7B,66.7),(3-4A,3-3B,33.3)
     (3-4A,3-4B,0.0),(3-4A,3-7B,66.7),(3-7A,3-3B,66.7),(3-7A,3-4B,66.7)
     (3-7A,3-7B,0.0)))
   ((4-4A,4-11B,75.0)
    ((3-1,3-2B,33.3),(3-1,3-4B,66.7),(3-1,3-6,66.7),(3-3A,3-2B,33.3)
     (3-3A,3-4B,33.3),(3-3A,3-6 66.7),(3-4A,3-2B,66.7),(3-4A,3-4B,0.0)
     (3-4A,3-6,66.7)))
   ((4-11A,4-2B,75.0)
    ((3-2A,3-1,33.3),(3-2A,3-2B,0.0),(3-2A,3-3B,33.3),(3-4A,3-1,66.7)
     (3-4A,3-2B,66.7),(3-4A,3-3B,33.3),(3-7B,3-1,66.7),(3-7B,3-2B,33.3)
     (3-7B,3-3B,66.7)))
   ((4-11A,4-4B,75.0)
    ((3-2A,3-1,33.3),(3-2A,3-3B,33.3),(3-2A,3-4B,66.7),(3-4A,3-1,66.7)
     (3-4A,3-3B,33.3),(3-4A,3-4B,0.0),(3-6,3-1,66.7),(3-6,3-3B,66.7)
     (3-6,3-4B,66.7)))
   ((4-11A,4-11B,75.0)
    ((3-2A,3-2B,0.0),(3-2A,3-4B,66.7),(3-2A,3-7A,33.3),(3-4A,3-2B,66.7)
     (3-4A,3-4B,0.0),(3-4A,3-7A,66.7),(3-7B,3-2B,33.3),(3-7B,3-4B,66.7)
     (3-7B,3-7A,0.0)))))
```

The upper-level values are updated with the lower-level ones. In all other respects Ex. 4.2.3 contains the same information as 4.2.2, but the SC names are omitted to make the hierarchy of values more obvious.

[6] When %REL_n is repeatedly evaluated within RECREL, values are represented as fractions and processed to full accuracy. Only the final RECREL result is rounded. In the examples depicting the comparison trees, the values are usually given to the accuracy of one decimal place.

EXAMPLE 4.2.3: RECREL(5-2A,5-2B), branch four. Values of the 86 %REL_n comparisons at levels four, three and two.

```
(60.0
  ((50.0 ( 0.0, 33.3, 33.3,  0.0,                                   ))
   (75.0 (33.3, 66.7, 33.3,  0.0, 33.3, 66.7, 66.7, 66.7, 66.7))
   (75.0 (33.3, 66.7, 66.7,  0.0, 66.7, 33.3, 33.3, 33.3, 66.7))
   (75.0 (33.3,  0.0, 66.7, 66.7, 33.3, 66.7, 33.3, 66.7, 66.7))
   (75.0 ( 0.0, 33.3, 66.7, 33.3,  0.0, 66.7, 66.7, 66.7,  0.0))
   (75.0 (33.3, 66.7, 66.7, 33.3, 33.3, 66.7, 66.7,  0.0, 66.7))
   (75.0 (33.3,  0.0, 33.3, 66.7, 66.7, 33.3, 66.7, 33.3, 66.7))
   (75.0 (33.3, 33.3, 66.7, 66.7, 33.3,  0.0, 66.7, 66.7, 66.7))
   (75.0 ( 0.0, 66.7, 33.3, 66.7,  0.0, 66.7, 33.3, 66.7,  0.0))))
```

The steps applied to update the values are as follows:

(1) Averages are taken from the values in the level-two cross-correlation groups.

EXAMPLE 4.2.4: RECREL(5-2A,5-2B), branch four. Values of comparisons at levels four and three, with averages values belonging to level-two cross-correlation groups.

```
(60.0
  ((50.0 (16.65))
   (75.0 (48.15))
   (75.0 (44.44))
   (75.0 (48.15))
   (75.0 (37.04))
   (75.0 (48.15))
   (75.0 (44.44))
   (75.0 (48.15))
   (75.0 (37.04))))
```

(2) Each level-three value is updated with the corresponding level-two average. The values are multiplied together and divided by 100.

EXAMPLE 4.2.5: RECREL(5-2A,5-2B), branch four. Updating level-three values with corresponding level-two averages.[7]

```
(60.0
  ((/ (* 50.0 16.65) 100))  = 8.325
   (/ (* 75.0 48.15) 100))  = 36.1125
   (/ (* 75.0 44.44) 100))  = 33.33
   (/ (* 75.0 48.15) 100))  = 36.1125
   (/ (* 75.0 37.04) 100))  = 27.78
   (/ (* 75.0 48.15) 100))  = 36.1125
   (/ (* 75.0 44.44) 100))  = 33.33
   (/ (* 75.0 48.15) 100))  = 36.1125
   (/ (* 75.0 37.04) 100))) = 27.78
```

(3) An average is taken from the updated level-three values.

[7] Order of functions and arguments as in the Common Lisp programming language.

EXAMPLE 4.2.6: RECREL(5-2A,5-2B), branch four. The average of the updated level-three values.

```
(60.0 (/ (+ 8.325 36.1125 33.33 36.1125 27.78 36.1125 33.33 36.1125 27.78) 9))
 = (60.0 (30.555))
```

(4) The final branch-four value, 18.3, is obtained through updating the one and only level-four value with the level-three average.

EXAMPLE 4.2.7: RECREL(5-2A,5-2B), branch four. The final value.

```
(/ (* 60.0 30.555) 100) = 18.3
```

The final branch-four value, 18.3, is considerably below the original top-level comparison value, 60. The tetrad class contents of the inversionally related pentad classes turned out to be more similar than the initial comparison suggested.

Branch three contains two levels, three and two. Value for the top-level comparison %REL$_3$(5-2A,5-2B) is 30. The 3C%Vs and difference vectors are given in Ex. 4.2.8.

EXAMPLE 4.2.8: The 3-class %-vectors of 5-2A (top row) and 5-2B (second row), a row of indexes and the difference vectors of 5-2A and 5-2B, respectively. %REL$_3$(5-2A,5-2B) = 30.

```
[20 20 10 10  0 10  0  0  0 10 10 10  0  0  0  0  0  0  0]
[20 10 20  0 10  0 10  0  0 10 10 10  0  0  0  0  0  0  0]

  1  2  2  3  3  4  4  5  5  6  7  7  8  8  9 10 11 11 12

[ 0 10  0 10  0 10  0  0  0  0  0  0  0  0  0  0  0  0  0]
[ 0  0 10  0 10  0 10  0  0  0  0  0  0  0  0  0  0  0  0]
```

The difference group of 5-2A contains the SCs {3-2A,3-3A,3-4A}, while that of 5-2B contains the SCs {3-2B,3-3B,3-4B}. The cross-correlation group, to be compared with %REL$_2$, is {{3-2A,3-2B}, {3-2A,3-3B}, {3-2A,3-4B}, {3-3A,3-2B}, {3-3A,3-3B},{3-3A,3-4B}, {3-4A,3-2B},{3-4A,3-3B},{3-4A,3-4B}}.

Example 4.2.9 gives the SC pairs and %REL$_n$ values of the ten comparisons in branch three. The one and only level-three comparison, between the pentad classes, is at the top. It is followed by the cross-correlation group derived from it.

EXAMPLE 4.2.9: RECREL(5-2A,5-2B), branch three. The ten $\%REL_n$ comparisons at levels three and two.

```
((5-2A,5-2B,30.0)
  ((3-2A,3-2B, 0.0),(3-2A,3-3B,33.3),(3-2A,3-4B,66.7)
   (3-3A,3-2B,33.3),(3-3A,3-3B, 0.0),(3-3A,3-4B,33.3)
   (3-4A,3-2B,66.7),(3-4A,3-3B,33.3),(3-4A,3-4B, 0.0)))
```

To obtain the final branch-three value, we first calculate the average of the nine level-two values. It is 29.6. Then, the one and only level-three value, 30, is updated with the average: (30 * 29.6)/100 = 8.88.

Calculating the value for branch two is easy. The only comparison to be made is $\%REL_2$(5-2A,5-2B). As the two inversionally related SCs have an identical 2C%V, [30 30 20 10 10 0], the $\%REL_2$ value is 0.

The value of branch four is 18.3, those of branches three and two 8.88 and 0, respectively. The final value is the average of these three values: RECREL(5-2A,5-2B) = (18.33 + 8.88 + 0)/3 ≈ 9.

4.4 FURTHER DETAILS

4.4.1 Weights

So far we have calculated a common value for a cross-correlation group simply by taking the average of the SC pair values in it. This is natural as long as we process cross-correlation groups like the one derived from the two difference vectors in Ex. 4.2.1. All difference vector components are the same, 20, suggesting that each subset-class with a unilaterally embedded share is of equal standing or importance. In fact, the nonzero components are uniform in all pairs of difference vectors produced during the entire RECREL(5-2A,5-2B) comparison. Here, taking an average is an accurate method of reducing several values into one.

We cannot expect to encounter such uniformity in every comparison, however, and must understand what sort of effects strong variances among difference vector components can have on cross-correlation group values. Ex. 4.3 gives two 3-class %-vectors that do not belong to any existing SCs. They were constructed only to give an extreme example for our discussion. For the sake of convenience, let us say that the top vector belongs to some hypothetical SC X, the lower one to some SC Y.

EXAMPLE 4.3: Two artificial 3C%Vs (rows 1 and 2), an index row and two difference vectors derived from the 3C%Vs. %REL3 value: 80.

```
[0 75  0  0  0  0  0  0  0  0  0  0  0  0  0  0  5  0 20]
[0  0 75  0  0  0  0  0  0  5  0  0  0  0  0  0  0  0 20]

 1  2  2  3  3  4  4  5  5  6  7  7  8  8  9 10 11 11 12

[0 75  0  0  0  0  0  0  0  0  0  0  0  0  0  0  5  0  0]
[0  0 75  0  0  0  0  0  0  5  0  0  0  0  0  0  0  0  0]
```

The difference group of X is {3-2A,3-11A}, that of Y {3-2B,3-6}. The cross-correlation group is {{3-2A,3-2B}, {3-2A,3-6}, {3-11A,3-2B}, {3-11A,3-6}}. The (rounded) %REL2 values of these pairs are 0, 67, 67 and 67, respectively. The average is 50.

In the upper difference vector, a major part of the triad class contents belong to 3-2A. 3-2B is equally well represented in the lower one. It is evident that the comparison %REL2(3-2A,3-2B) will have an exceptional prominence among the comparisons, regardless of the actual value it produces. In other words, the degree of similarity between the unilaterally embedded subset-class materials in X and Y is to a very high extent compatible with the degree of similarity between 3-2A and 3-2B, whatever that degree is. Likewise, when we compare 3-11A and 3-6, the small components indicate that the resulting value, 67, illustrates the degree of similarity existing within only a fraction of the unilaterally embedded subset-class materials.

Obviously, then, there are two independent aspects to every comparison, the %RELn value itself and the *weight* the comparison has among all comparisons in a cross-correlation group. The latter is to be calculated from the two individual weights each compared pair has, the individual weights themselves being derived from difference vector components. Before a component can become the weight of a subset-class, however, it must be slightly modified.

4.4.2 Scaled Difference Vectors

In Ex. 4.3 the %REL3(X,Y) value 80 indicates (indirectly) that 20% of the proportioned triad class contents of X and Y are the same. These portions are excluded from further comparisons, leaving the focus solely on the unilaterally embedded subset-class shares. In spite of this, the mutually embedded shares still have an effect on the two difference vectors: in both of them, the sum of components is 80. As the subset-classes having unilaterally embedded shares now constitute a full 100% of the material we are examining, also their weights must be able to cover 100% of it. The difference vectors must be scaled so that each component is divided by the sum of the components and multiplied by 100. The sum of components in each resulting *scaled difference vector* is 100. Ex. 4.4 gives the scaled difference vectors derived from the difference vectors in Ex. 4.3.

EXAMPLE 4.4: Two scaled difference vectors derived from the difference vectors in Ex. 4.3.

```
[0 94  0  0  0  0  0  0  0  0  0  0  0  0  0  0  6  0  0]
[0  0 94  0  0  0  0  0  0  6  0  0  0  0  0  0  0  0  0]
 1  2  2  3  3  4  4  5  5  6  7  7  8  8  9 10 11 11 12
```

4.4.3 Weights in Difference Groups and Cross-Correlation Groups

From now on, each subset-class in a difference group is to be associated with the weight its scaled difference vector component indicates. By convention we will give the weight first, then the SC name. In Ex. 4.3 the difference group of X is {(94,3-2A), (6,3-11A)} and that of Y {(94,3-2B),(6,3-6)}.

Accordingly, when a cross-correlation group is derived from two difference groups, each SC pair is to be associated with the weights of the classes. The cross-correlation group derived from the X- and Y-difference groups above is {(94,94,3-2A,3-2B), (94,6,3-2A,3-6), (6,94,3-11A,3-2B), (6,6,3-11A,3-6)}. Each cell gives the weight of the X-class, the weight of the Y-class, the X-class and the Y-class.

4.4.4 Processing a Cross-Correlation Group

A single value representing an entire cross-correlation group is derived from all $\%REL_n$ values and weights in the group. To do this, two new concepts are needed.

4.4.4.1 Proportioned Weights

First, a combined weight is determined for each pair from the two class weights. Given weights w_1 and w_2, the *proportioned weight* $w_p = (w_1 * w_2)/100$. It gives the percentual weight of the SC pair among all pairs in the cross-correlation group. The sum of proportioned weights in all cross-correlation groups is 100.[8]

[8] Each comparison within a given branch gets its proportioned weight in due course, with the exception of the very first one. The two top-level SCs do not "inherit" weights from previous comparisons. In order to make the process of updating the branch values consistent, we determine weights and a proportioned weight also for the top-level pair. As the pair never shares the top level with any other

As an example, let us process the cross-correlation group in section 4.4.3 so that a proportioned weight is calculated from each weight pair. Each element in the modified cross-correlation group now contains the proportioned weight, the SCs and the $\%REL_2$ value of the pair: {(88.4,3-2A,3-2B,0), (5.6,3-2A,3-6,67), (5.6,3-11A,3-2B,67), (0.4,3-11A,3-6,67)}.

The proportioned weights show, in concrete numbers, how taking an average from cross-correlation group values can grossly under-represent some comparisons and over-represent others. In our artificial example, the first pair covers nearly 90% of the unilaterally embedded subset-class materials, the last one less than 1%.

4.4.4.2 Weighted Values

Each $\%REL_n$ value is related to its corresponding proportioned weight, the result reflecting both the degree of similarity between the SCs and the importance of the comparison within the cross-correlation group.

Given a $\%REL_n$ value v and a proportioned weight w_p, the resulting *weighted value* $v_w = (v * w_p)/100$. Modified so that each set-class pair is followed by its weighted value, the cross-correlation group in section 4.4.4.1 above is as follows: {(3-2A,3-2B,0), (3-2A,3-6,3.75), (3-11A,3-2B,3.75), (3-11A,3-6,0.3)}.

4.4.4.3 Deriving the Value Representing a Cross-Correlation Group

The value representing an entire cross-correlation group is the sum of all weighted values in it. This value is the *weighted arithmetic mean* of the values. In the cross-correlation group of section 4.4.4.2 above, the sum is (0 + 3.75 + 3.75 + 0.3) = 7.8. The weighted arithmetic mean is substantially lower than the average value 50 which we calculated for this same cross-correlation group in section 4.4.1.

4.4.4.4 Updating Weighted Values

The updating process described in sections 4.2 and 4.3.2 is practically identical to the one we will use with our new and final method for determining a common value for a cross-correlation group. Predictably, the only difference is that while in the initial

pairs, both SCs get the weight 100. Consequently, the proportioned weight is always 100 as well.

version we calculated the averages of the cross-correlation group values, from now on we calculate the sums of weighted values.

Given, within some branch, the level-two cross-correlation groups C_1, C_2,..C_n and the sums s_1, s_2,..s_n of weighted values in them, respectively, each sum s_i updates the weighted value v_W of the level-three SC pair P_i from which C_i was derived: $(s_i * v_W)/100$. The result is interpreted as the new weighted value of P_i. This SC, in turn, belongs to one of the level-three cross-correlation groups C_1, C_2,..C_n. The sums s_1, s_2, ..s_n of (updated) weighted values are taken in each of them. Each sum s_i updates the weighted value of the level-four SC pair from which C_i was derived, etc., until the weighted value of the top-level pair has been updated with the weighted arithmetic mean of the only cross-correlation group derived from the pair.

Concrete examples of the updating process will be given during the RECREL comparison below.

4.5 AN EXAMPLE: RECREL(7-35,4-22A)

The whole RECREL mechanism has now been introduced and analysed. In the following we will once more go through an entire comparison, describing the different phases in it in relatively close detail. The classes to be compared are 7-35, {0,1,3,5,6,8,10}, the "diatonic class," and one of its 4-pc subset-classes, 4-22A, {0,2,4,7}. As the SCs are of different cardinalities, the highest branch equals the smaller cardinality. The branches are four, three and two.

4.5.1 Branch Four

Within branch four, the levels are four, three and two. The only level-four comparison, %REL_4(7-35,4-22A), has the value 91. The value is predictably high, as 4-22A is itself the only SC with a nonzero component in 4C%V(4-22A). The 4C%Vs, difference vectors and scaled difference vectors of 7-35 and 4-22A are given in example 4.5.1.

The difference groups will be derived from the lowest pair of vectors. The 19 weight/SC pairs in the 7-35 group are {(3,4-8),(6,4-10),(6,4-11A),(6,4-11B), etc. The 4-22A group contains only one weight/SC pair, {(100,4-22A)}. In the level-three cross-correlation group to be derived from these two difference groups, 4-22A participates in each of the 19 SC pairs. With the weight and SC pairs, the group elements are {(3,100,4-8,4-22A), (6,100,4-10,4-22A), (6,100,4-11A,4-22A), etc.

EXAMPLE 4.5.1: The 4-class %-vectors, difference vectors and scaled difference vectors of 7-35 and 4-22A, with two rows of indexes. In each vector pair, the upper one belongs to 7-35. %REL4(7-35,4-22A) = 91. Component C = 100.

```
[0 0 0 0 0 0 0 0 0  0 3 0 6 6 6 0 0 3 3 6 6  0 0 3 3 0 0 0 0 0 6 3 9 9 11 0 0 9  3 3 0 3  3]
[0 0 0 0 0 0 0 0 0  0 0 0 0 0 0 0 0 0 0 0 0  0 0 0 0 0 0 0 0 0 0 C 0 0 0  0 0 0 0 0 0 0 0]

 1 2 2 3 4 4 5 5 6 7 8 9 10 11 11 12 12 13 13 14 14 15 15 16 16 17 18 18 19 19 20 21 22 22 23 24 25 26 27 27 28 29 29

[0 0 0 0 0 0 0 0 0 0 3 0 6 6  6 0 0 3 3 6 6 0 0 3 3  0 0 0 0 0  6 3  0 9 11  0 0 9 3 3 0 3 3]
[0 0 0 0 0 0 0 0 0 0 0 0 0 0  0 0 0 0 0 0 0 0 0 0 0  0 0 0 0 0  0 0 91 0 0  0 0 0 0 0 0 0 0]

 1 2 2 3 4 4 5 5 6 7 8 9 10 11 11 12 12 13 13 14 14 15 15 16 16 17 18 18 19 19 20 21 22 22 23 24 25 26 27 27 28 29 29

[0 0 0 0 0  0 0 0 0 0 3 0 6 6 6 0 0  3 3 6 6 0 0  3 3 0  0 0 0 0 6 3 0 9 12  0 0 9 3 3 0 3 3]
[0 0 0 0 0  0 0 0 0 0 0 0 0 0 0 0 0  0 0 0 0 0 0  0 0 0  0 0 0 0 0 0 C 0  0 0 0 0 0 0 0 0 0]
```

Each tetrad class pair is compared with %REL3, and a level-two cross-correlation group is derived from it. As an example we examine the pair {4-11B,4-22}. The three types of vector pairs are given in Ex. 4.5.2.

%REL3(4-11B,4-22A) = 50. The difference group belonging to 4-11B is {(50,3-2B), (50,3-4B)}. The difference group of 4-22A, in turn, is {(50,3-9), (50,3-11B)}. The resulting 4-element cross-correlation group contains weight pair / SC pair elements {(50,50,3-2B,3-9), (50,50,3-2B,3-11B), (50,50,3-4B,3-9), (50,50,3-4B,3-11B)}.

EXAMPLE 4.5.2: The 3-class %-vectors, difference vectors and scaled difference vectors of SCs 4-11B and 4-22A, with two rows of indexes. In each vector pair, the upper one belongs to 4-11B. %REL3(4-11B,4-22A) = 50.

```
[0   0 25   0   0   0 25   0   0 25 25   0   0   0   0   0   0   0   0]
[0   0  0   0   0   0  0   0   0 25 25   0   0   0 25   0   0  25   0]

 1   2  2   3   3   4  4   5   5  6   7   7   8   8   9  10  11  11  12

[0   0 25   0   0   0 25   0   0  0   0   0   0   0   0   0   0   0   0]
[0   0  0   0   0   0  0   0   0  0   0   0   0   0  25   0   0  25   0]

 1   2  2   3   3   4  4   5   5  6   7   7   8   8   9  10  11  11  12

[0   0 50   0   0   0 50   0   0  0   0   0   0   0   0   0   0   0   0]
[0   0  0   0   0   0  0   0   0  0   0   0   0   0  50   0   0  50   0]
```

In the 19 level-two cross-correlation groups, the total number of triad class pairs is 167. Each pair gets a %REL2 value and from the two class weights, a proportioned weight. Similarly, the level-three pairs get their %REL3 values and proportioned weights. The top-level pair, {7-35,4-22A}, has 91 as its %REL4 value, and as it does not share level four with any other pairs, its proportioned weight is 100.

EXAMPLE 4.5.3: RECREL(7-35,4-22A). Branch four, part of the comparisons at levels 4 - 2. Each element contains a proportioned weight, a SC pair and the $\%REL_n$ value of the pair.

```
(100,7-35,4-22A,91.4)
  (3.1,4-8,4-22A,100.0)
     (6.2,3-4A,3-6,66.7),(6.2,3-4A,3-7A,66.7),(6.2,3-4A,3-9,66.7),(6.2,3-4A,3-11B,33.3)
     (6.2,3-4B,3-6,66.7),(6.2,3-4B,3-7A,66.7),(6.2,3-4B,3-9,66.7),(6.2,3-4B,3-11B,33.3)
     (6.2,3-5A,3-6,100.0),(6.2,3-5A,3-7A,66.7),(6.2,3-5A,3-9,66.7),(6.2,3-5A,3-11B,66.7)
     (6.2,3-5B,3-6,100.0),(6.2,3-5B,3-7A,66.7),(6.2,3-5B,3-9,66.7),(6.2,3-5B,3-11B,66.7)
  (6.2,4-10,4-22A,75.0)
     (11.1,3-2A,3-6,66.7),(11.1,3-2A,3-9,66.7),(11.1,3-2A,3-11B,66.7),(11.1,3-2B,3-6,66.7)
     (11.1,3-2B,3-9,66.7),(11.1,3-2B,3-11B,66.7),(11.1,3-7B,3-6,66.7),(11.1,3-7B,3-9,33.3)
     (11.1,3-7B,3-11B,33.3)
  (6.2,4-11A,4-22A,75.0)
     (11.1,3-2A,3-7A,33.3),(11.1,3-2A,3-9,66.7),(11.1,3-2A,3-11B,66.7),(11.1,3-4A,3-7A,66.7)
     (11.1,3-4A,3-9,66.7),(11.1,3-4A,3-11B,33.3),(11.1,3-7B,3-7A,0.0),(11.1,3-7B,3-9,33.3)
     (11.1,3-7B,3-11B,33.3)
  (6.2,4-11B,4-22A,50.0)
     (25.0,3-2B,3-9,66.7),(25.0,3-2B,3-11B,66.7),(25.0,3-4B,3-9,66.7),(25.0,3-4B,3-11B,33.3)
  (3.1,4-13A,4-22A,75.0)
     (11.1,3-2A,3-6,66.7),(11.1,3-2A,3-9,66.7),(11.1,3-2A,3-11B,66.7),(11.1,3-5A,3-6,100.0)
     (11.1,3-5A,3-9,66.7),(11.1,3-5A,3-11B,66.7),(11.1,3-10,3-6,100.0),(11.1,3-10,3-9,100.0)
     (11.1,3-10,3-11B,66.7)........
```

The total number of $\%REL_n$ comparisons in branch four is 187. Ex. 4.5.3 gives a part of the comparisons as a diagram. Each element contains a proportioned weight, a SC pair and a $\%REL_n$ value. Level structure can be inferred from the indentation.

A weighted value is calculated for each SC pair from its proportioned weight and $\%REL_n$ value. Ex. 4.5.4 shows a part of the branch-four comparisons after this task has been completed.

EXAMPLE 4.5.4: RECREL(7-35,4-22A). Branch four, part of the comparisons at levels 4 - 2. Each element contains a SC pair and a weighted value.

```
(7-35,4-22A,91.4)
  (4-8,4-22A,3.1)
     (3-4A,3-6,4.2),(3-4A,3-7A,4.2),(3-4A,3-9,4.2),(3-4A,3-11B,2.1),(3-4B,3-6,4.2)
     (3-4B,3-7A,4.2),(3-4B,3-9,4.2),(3-4B,3-11B,2.1),(3-5A,3-6,6.2),(3-5A,3-7A,4.2)
     (3-5A,3-9,4.2),(3-5A,3-11B,4.2),(3-5B,3-6,6.2),(3-5B,3-7A,4.2),(3-5B,3-9,4.2)
     (3-5B,3-11B,4.2)
  (4-10,4-22A,4.7)
     (3-2A,3-6,7.4),(3-2A,3-9,7.4),(3-2A,3-11B,7.4),(3-2B,3-6,7.4),(3-2B,3-9,7.4)
     (3-2B,3-11B,7.4),(3-7B,3-6,7.4),(3-7B,3-9,3.7),(3-7B,3-11B,3.7)
  (4-11A,4-22A,4.7)
     (3-2A,3-7A,3.7),(3-2A,3-9,7.4),(3-2A,3-11B,7.4),(3-4A,3-7A,7.4),(3-4A,3-9,7.4)
     (3-4A,3-11B,3.7),(3-7B,3-7A,0.0),(3-7B,3-9,3.7),(3-7B,3-11B,3.7)
  (4-11B,4-22A,3.1),
     (3-2B,3-9,16.7),(3-2B,3-11B,16.7),(3-4B,3-9,16.7),(3-4B,3-11B,8.3)
  (4-13A,4-22A,2.3)
     (3-2A,3-6,7.4),(3-2A,3-9,7.4),(3-2A,3-11B,7.4),(3-5A,3-6,11.1),(3-5A,3-9,7.4)
     (3-5A,3-11B,7.4),(3-10,3-6,11.1),(3-10,3-9,11.1),(3-10,3-11B,7.4).......
```

EXAMPLE 4.5.5: RECREL(7-35,4-22A). Branch four, values of all comparisons at levels 4 - 2. Updating the values.

```
(91.4
 (/ (* 3.1 (+ 4.2 4.2 4.2 2.1 4.2 4.2 4.2 2.1 6.2 4.2 4.2 4.2 6.2 4.2 4.2 4.2)) 100)= 2.07
 (/ (* 4.7 (+ 7.4 7.4 7.4 7.4 7.4 7.4 7.4 3.7 3.7)) 100)                          = 2.78
 (/ (* 4.7 (+ 3.7 7.4 7.4 7.4 7.4 3.7 0.0 3.7 3.7)) 100)                          = 2.08
 (/ (* 3.1 (+ 16.7 16.7 16.7 8.3)) 100)                                           = 1.81
 (/ (* 2.3 (+ 7.4 7.4 7.4 11.1 7.4 7.4 11.1 11.1 7.4)) 100)                       = 1.78
 (/ (* 3.1 (+ 4.2 2.1 4.2 4.2 6.2 4.2 4.2 4.2 4.2 0.0 2.1 2.1 6.2 4.2 6.2 4.2)) 100)= 1.94
 (/ (* 4.7 (+ 7.4 3.7 7.4 7.4 7.4 3.7 7.4 3.7 0.0)) 100)                          = 2.26
 (/ (* 3.1 (+ 16.7 8.3 16.7 16.7)) 100)                                           = 1.81
 (/ (* 2.3 (+ 7.4 7.4 3.7 11.1 7.4 7.4 3.7 7.4 7.4)) 100)                         = 1.44
 (/ (* 2.3 (+ 7.4 7.4 3.7 11.1 7.4 7.4 3.7 7.4 7.4)) 100)                         = 1.44
 (/ (* 4.7 (+ 7.4 7.4 7.4 7.4 7.4 7.4 7.4 3.7 7.4)) 100)                          = 2.95
 (/ (* 2.3 (+ 7.4 7.4 7.4 7.4 7.4 7.4 7.4 7.4 7.4)) 100)                          = 1.53
 (/ (* 4.7 (+ 0.0 8.3 8.3 0.0)) 100)                                              = 0.78
 (/ (* 6.2 (+ 16.7 8.3 16.7 16.7)) 100)                                           = 3.62
 (/ (* 4.7 (+ 16.7 8.3 16.7 16.7)) 100)                                           = 2.74
 (/ (* 2.3 (+ 3.7 7.4 7.4 11.1 11.1 7.4 7.4 7.4 0.0)) 100)                        = 1.44
 (/ (* 2.3 (+ 7.4 0.0 3.7 3.7 7.4 7.4 11.1 7.4 11.1)) 100)                        = 1.36
 (/ (* 3.1 (+ 4.2 2.1 4.2 4.2 6.2 4.2 4.2 4.2 2.1 4.2 4.2 4.2 4.2 2.1 4.2 0.0)) 100)= 1.81
 (/ (* 2.3 (+ 7.4 3.7 7.4 11.1 7.4 7.4 3.7 7.4 7.4)) 100)                         = 1.44

  = (/ (* 91.4 (+ 2.07 2.78 2.08 1.81 1.78 1.94 2.26 1.81 1.44 1.44 2.95 1.53 0.78 3.62
                2.74 1.44 1.36 1.81 1.44)) 100)

  = (/ (* 91.4 37.08) 100) = 33.89
```

The updating is begun by taking the sums of weighted values in the level-two cross-correlation groups and updating the level-three weighted values with them. The updated level-three weighted values are added together, the sum updating the only level-four weighted value. These stages are described in Ex. 4.5.5. The SC names are omitted.

The branch-four value is 33.89.

4.5.2 Branch Three

The levels within branch three are three and two. The only level-three comparison is %REL3(7-35,4-22A), having the value 57. The 3C%Vs, difference vectors and scaled difference vectors of the two classes are in Ex. 4.5.6.

EXAMPLE 4.5.6: The 3-class %-vectors, difference vectors and scaled difference vectors of 7-35 and 4-22A, with two rows of indexes. In each vector pair, the upper belongs to 7-35. $\%REL_3$(7-35,4-22A) = 57.

```
[0  6  6  0  0  6  6  3  3  9 11 11  3  3 14  3  9  9  0]
[0  0  0  0  0  0  0  0  0 25 25  0  0  0 25  0  0 25  0]

 1  2  2  3  3  4  4  5  5  6  7  7  8  8  9 10 11 11 12

[0  6  6  0  0  6  6  3  3  0  0 11  3  3  0  3  9  0  0]
[0  0  0  0  0  0  0  0  0 16 14  0  0  0 11  0  0 16  0]

 1  2  2  3  3  4  4  5  5  6  7  7  8  8  9 10 11 11 12

[0 10 10  0  0 10 10  5  5  0  0 20  5  5  0  5 15  0  0]
[0  0  0  0  0  0  0  0  0 29 24  0  0  0 19  0  0 29  0]
```

The nonzero scaled difference vector components and their indexes again give the difference groups. These are {(10,3-2A), (10,3-2B), (10,3-4A), etc., for 7-35, {(29,3-6), (24,3-7A), (19,3-9), (29,3-11B)} for 4-22A. The resulting level-two cross-correlation group contains 44 weight pair / SC pair elements, starting with {(10,29,3-2A,3-6), (10,24,3-2A,3-7A), (10,19,3-2A,3-9), etc.

EXAMPLE 4.5.7: RECREL(7-35,4-22A). Branch three, the comparisons at levels 3 and 2. Each element contains a proportioned weight, a SC pair and a $\%REL_n$ value.

```
((100,7-35,4-22A,57.1)
 ((2.9,3-2A,3-6,66.7),(2.4,3-2A,3-7A,33.3),(1.9,3-2A,3-9,66.7),(2.9,3-2A,3-11B,66.7)
  (2.9,3-2B,3-6,66.7),(2.4,3-2B,3-7A,33.3),(1.9,3-2B,3-9,66.7),(2.9,3-2B,3-11B,66.7)
  (2.9,3-4A,3-6,66.7),(2.4,3-4A,3-7A,66.7),(1.9,3-4A,3-9,66.7),(2.9,3-4A,3-11B,33.3)
  (2.9,3-4B,3-6,66.7),(2.4,3-4B,3-7A,66.7),(1.9,3-4B,3-9,66.7),(2.9,3-4B,3-11B,33.3)
  (1.4,3-5A,3-6,100.0),(1.2,3-5A,3-7A,66.7),(0.9,3-5A,3-9,66.7),(1.4,3-5A,3-11B,66.7)
  (1.4,3-5B,3-6,100.0),(1.2,3-5B,3-7A,66.7),(0.9,3-5B,3-9,66.7),(1.4,3-5B,3-11B,66.7)
  (5.7,3-7B,3-6,66.7),(4.7,3-7B,3-7A,0.0),(3.7,3-7B,3-9,33.3),(5.7,3-7B,3-11B,33.3)
  (1.4,3-8A,3-6,33.3),(1.2,3-8A,3-7A,66.7),(0.9,3-8A,3-9,66.7),(1.4,3-8A,3-11B,66.7)
  (1.4,3-8B,3-6,33.3),(1.2,3-8B,3-7A,66.7),(0.9,3-8B,3-9,66.7),(1.4,3-8B,3-11B,66.7)
  (1.4,3-10,3-6,100.0),(1.2,3-10,3-7A,66.7),(0.9,3-10,3-9,100.0),(1.4,3-10,3-11B,66.7)
  (4.3,3-11A,3-6,66.7),(3.6,3-11A,3-7A,33.3),(2.8,3-11A,3-9,66.7),(4.3,3-11A,3-11B,0.0)))
```

All 45 branch-three comparisons are given in Ex. 4.5.7. In the example, the proportioned weights have already been calculated and are given with the SC names and $\%REL_n$ values. The top-level comparison again receives the proportioned weight 100.

A weighted value is calculated for each SC pair from its proportioned weight and $\%REL_n$ value. Ex. 4.5.8 gives the branch-three comparisons after this phase has been completed.

EXAMPLE 4.5.8: RECREL(7-35,4-22A). Branch three, the comparisons at levels 3 and 2. Each element contains a SC pair and a weighted value.

```
((7-35,4-22A,57.1)
  (3-2A,3-6,1.9),(3-2A,3-7A,0.8),(3-2A,3-9,1.2),(3-2A,3-11B,1.9),(3-2B,3-6,1.9),
  (3-2B,3-7A,0.8),(3-2B,3-9,1.2),(3-2B,3-11B,1.9),(3-4A,3-6,1.9),(3-4A,3-7A,1.6),
  (3-4A,3-9,1.2),(3-4A,3-11B,1.0),(3-4B,3-6,1.9),(3-4B,3-7A,1.6),(3-4B,3-9,1.2),
  (3-4B,3-11B,1.0),(3-5A,3-6,1.4),(3-5A,3-7A,0.8),(3-5A,3-9,0.6),(3-5A,3-11B,1.0),
  (3-5B,3-6,1.4),(3-5B,3-7A,0.8),(3-5B,3-9,0.6),(3-5B,3-11B,1.0),(3-7B,3-6,3.8),
  (3-7B,3-7A,0.0),(3-7B,3-9,1.2),(3-7B,3-11B,1.9),(3-8A,3-6,0.5),(3-8A,3-7A,0.8),
  (3-8A,3-9,0.6),(3-8A,3-11B,1.0),(3-8B,3-6,0.5),(3-8B,3-7A,0.8),(3-8B,3-9,0.6),
  (3-8B,3-11B,1.0),(3-10,3-6,1.4),(3-10,3-7A,0.8),(3-10,3-9,0.9),(3-10,3-11B,1.0),
  (3-11A,3-6,2.9)(3-11A,3-7A,1.2),(3-11A,3-9,1.9),(3-11A,3-11B,0.0))))
```

Updating is less complicated than in branch four, as the number of levels is only two. All 44 level-two weighted values are added together, the sum updating the one and only level-three weighted value. The final branch-three value is 30.49. These steps are described in Ex. 4.5.9. The SC names are omitted.

EXAMPLE 4.5.9: RECREL(7-35,4-22A). Branch three, the comparisons at levels 3 and 2. Updating the values.

```
(/ (* 57.1
      (+ 1.9 0.8 1.2 1.9 1.9 0.8 1.2 1.9 1.9 1.6 1.2 1.0 1.9 1.6 1.2 1.0 1.4 0.8 0.6
         1.0 1.4 0.8 0.6 1.0 3.8 0.0 1.2 1.9 0.5 0.8 0.6 1.0 0.5 0.8 0.6 1.0 1.4 0.8
         0.9 1.0 2.9 1.2 1.9 0.0)) 100)

= (/ (* 57.1  53.39) 100) = 30.49
```

4.5.3 Branch Two

Branch two consists of only one comparison, $\%REL_2$(7-35,4-22A). The value is 17. The two dyad class %-vectors and the absolute values of the component pair differences are given in Ex. 4.5.10.

EXAMPLE 4.5.10: The dyad class %-vectors of 7-35 (top) and 4-22A. $\%REL_2$(7-35,4-22A) = 17.

```
  [10 24 19 14 29  5]
  [ 0 33 17 17 33  0]
 +|10  9  2  3  4  5|/2 ≈ 17
```

4.5.4 The Final RECREL Value

To get the final value of the entire comparison, the branch values are added together and divided by three. RECREL(7-35,4-22A) = (33.89 + 30.49 + 17)/3 ≈ 27. To reach this value, $\%REL_n$ was evaluated 233 times.

4.6 RECREL FORMALISATIONS

The following section 4.6.1 contains a verbal algorithm, and in section 4.6.2 the algorithm is converted to a recursive Common Lisp function.

4.6.1 An Algorithm: RECREL in Ten Phases

(1) Select the SCs {X,Y} to be compared. Determine the highest branch *branch*. If *branch* = 2, RECREL(X,Y) = $\%REL_2$(X,Y). Otherwise, within branch *branch*, set the highest level *level* = *branch*. In each branch, set the proportioned weight w_p of {X,Y} to 100.

(2) Set the variable {SC_1,SC_2} to {X,Y}. Set the variable b to *branch*. Set the variable n to *level*.

(3) Associate {SC_1,SC_2} with its weighted value $v_w = (v * w_p)/100$. $v = \%REL_n(SC_1,SC_2)$. w_p = the proportioned weight of {SC_1,SC_2}.

(4) Derive scaled difference vectors from nC%V(SC_1) and nC%V(SC_2). Collect the nonzero components and their indexes from the scaled difference vectors. Identify the SCs to which the indexes refer. Interpret the two groups of component-index pairs to be two difference groups, where each weight w is associated with a SC. Make a cross-correlation group. Replace the two weights w_1 and w_2 in each pair with the proportioned weight $w_p = (w_1 * w_2)/100$.

(5) Set new n by decreasing the current n by 1. Obtain one or more cross-correlation groups formed during the previous phase. In turn, set each SC pair in the group or groups to be {SC_1,SC_2}. If $n > 2$, take each {SC_1,SC_2} through phases 3 and 4. If $n = 2$, take each {SC_1,SC_2} through phase 3 only and go then straight to 6. Retain the cross-correlation group composition, knowing which pairs constitute a given group.

(6) Collect all cross-correlation groups C_1, C_2,..C_m whose SC pairs were compared with the current-n $\%REL_n$. Associate each cross-correlation group C_i with the sum s_i of weighted values in it. Take the cross-correlation group C_1 and seek the SC pair P_1 that was compared with $\%REL_{n+1}$ at an earlier stage and from which C_1 was derived. Obtain the weighted value v_w of P_1. Update v_w with the C_1 sum s_1 by calcu-

lating $(s_1 * v_w)/100$. Consider the updated value to be the new weighted value v_w of P_1. Repeat this for the rest of the cross-correlation groups $C_2,..C_m$ and the corresponding SC pairs $P_2,..P_m$ from which the groups were derived.

If (b-n) = 1, consider the updated weighted value of the one and only SC pair P = {X,Y} to be the final value of the branch b, and go straight to 8. Otherwise, go to 7.

(7) Set new n by increasing the current one by 1. Go to 6.

(8) Set new b by decreasing the current one by 1. If b is now 2, go straight to 10. Otherwise, set n to b and go to 9.

(9) Set $\{SC_1,SC_2\}$ to {X,Y}. Go to 3.

(10) Calculate branch 2 by evaluating $\%REL_2(X,Y)$. Calculate the final RECREL value by taking the average of the individual branch values.

4.6.2 RECREL as a Common Lisp Function

For programming convenience, the function may execute some internal phases of a RECREL comparison in an order that differs slightly from the one given in the algorithm above. This has no effect on the results. The numbers in parentheses preceded by semicolons refer to comments below.

The function is intended to be as complete as possible, performing independently almost all phases of a RECREL comparison.[9] Two external non-standard Common Lisp functions are evoked during its evaluation, however. Incorporating them would not have been of benefit from the point of view of clarity of the code. These are (a) the function *cardinality?*, returning the cardinality of the argument SC, (b) the function *cardinality-class*, returning all SCs in the cardinality-class n, $2 \leq n \leq 12$.

Furthermore, the %-vectors needed during the comparisons are stored in a table named **%-vec-hash**. To retrieve the n-class %-vector of SC X, the expression *(nth (- n 2) (gethash X *%-vec-hash*))* is evaluated.

[9] The function is made for demonstrational purposes only. Using it to perform RECREL comparisons between SCs of large cardinalities is impossible without a computer having considerable computing and memory capacities. The entire set of RECREL values was originally calculated with the help of a program using tables of precalculated branch values.

```
(defun RECREL (SC1 SC2)                                                  ; (1)
  (let* ((card1 (cardinality? SC1)) (card2 (cardinality? SC2))           ; (2)
         (branch (if (/= card1 card2) (min card1 card2) (1- card1)))
         (start1 (list (list 100 SC1))) (start2 (list (list 100 SC2))) B-v-l)
    (if (>= 2 branch)                                                    ; (3)
      (round (/ (apply #'+ (mapcar #'(lambda (num1 num2) (abs (- num1 num2)))
                               (nth 0 (gethash SC1 *%-vec-hash*))
                               (nth 0 (gethash SC2 *%-vec-hash*)))) 2))
      (loop                                                              ; (4)
        (if (= 1 branch)                                                 ; (8)
          (return (round (/ (apply #'+ B-v-l) (length B-v-l))))          ; (9)
          (progn
            (push
             (labels
               ((levels (w+SC-ls1 w+SC-ls2 level)                        ; (5)
                  (let ((B-v 0) vec1 vec2 (card-cl (cardinality-class level)) ; (5.2)
                        (i 0) dnow DG1 DG2  (%-v 0))
                    (dolist (w+SC1 w+SC-ls1 B-v)                         ; (5.6)
                      (setf vec1 (nth (- level 2)
                                      (gethash (second w+SC1) *%-vec-hash*)))
                      (dolist (w+SC2 w+SC-ls2)
                        (setf vec2 (nth (- level 2)
                                        (gethash (second w+SC2) *%-vec-hash*)))
                        (dolist
                          (item vec1 (if (= 2 level)
                                       (incf B-v                         ; (5.8)
                                         (/ (* (/ (* (first w+SC1) (first w+SC2))
                                                  100) %-v) 100))
                                       (incf B-v
                                         (/ (*
                                             (/ (* (/ (* (first w+SC1)
                                                         (first w+SC2)) 100) %-v) 100)
                                             (levels                     ; (5.7)
                                              (mapcar #'(lambda (pair)   ; (5.5)
                                                          (cons (* (/ (first pair) %-v) 100)
                                                                (rest pair))) DG1)
                                              (mapcar #'(lambda (pair)
                                                          (cons (* (/ (first pair) %-v) 100)
                                                                (rest pair))) DG2)
                                              (1- level))) 100))))
                          (setf dnow (- item (pop vec2)))
                          (cond
                           ((plusp dnow) (incf %-v dnow)                 ; (5.3)
                            (push (list dnow (nth i card-cl)) DG1) (incf i)) ; (5.4)
                           ((minusp dnow)
                            (push (list (- dnow) (nth i card-cl)) DG2) (incf i)) ; (5.4)
                           (t (incf i))))
                        (setf DG1 nil DG2 nil i 0 %-v 0))))))
               (levels start1 start2 branch))                            ; (5.1)
             B-v-l)                                                      ; (6)
            (setf branch (1- branch))))))))                              ; (7)
```

The comments

(1) The argument list is given here, containing two SCs *SC1* and *SC2*. (2) The cardinalities of the SCs are found out and the highest branch *branch* determined. The weight 100 is associated with *SC1* and *SC2*, resulting in *start1* = ((100 *SC1*)) and *start2* = ((100 *SC2*)). (3) If *branch* = 2, the %REL_2 value between *SC1* and *SC2* is established and returned as the final RECREL value. (4) If *branch* > 2, the evaluation of the function continues.

(5) *levels* is a recursive local function within RECREL, returning branch val-

ues. Its arguments *w+SC-ls1* and *w+SC-ls2* are difference groups with other names, containing one or more weight/SC pairs.[10] The third argument, *level*, takes its value from the current level. (5.1) Before the processing of a given branch is begun with *levels*, the arguments defining the top-level comparison are fetched from here. The first *w+SC-ls1* = *start1* = *((100 SC1))* and the first *w+SC-ls2* = *start2* = *((100 SC2))*. *level* = *branch*. (5.2) Among the *levels* variables: *B-v*, for *Branch value; card-cl*, for *cardinality class*, for identifying the SC names of the difference group classes; *DG1* and *DG2*, for *Difference Group 1* and *2; %-v*, for *%REL*$_n$ *value*.

During its various stages, *levels* (5.3) calculates %REL$_n$ values, (5.4) gathers weight/SC pairs from %-vectors, (5.5) scales the weights to produce proper difference groups, (5.6) forms cross-correlations groups, (5.7) calls itself with new arguments to perform lower-level comparisons and derive difference groups, cross-correlation groups, etc., from these, (5.8) calculates proportioned weights and weighted values, updating the latter until a final branch value is reached.

(6) When the value of a given branch is found out, it is stored in *B-v-l*, for *branch value list*. (7) *branch* is decreased by one. (8) *branch* is tested. If it is higher than 1, the comparison continues from *Levels*. If it is 1, (9) the final value for RECREL(*SC1,SC2*) is established by taking the sum of the *B-v-l* values, dividing it by the length of the *B-v-l* list and rounding the quotient.

4.7 RECREL AND THE EVALUATION CRITERIA

RECREL meets all criteria defined in section 2.3.

[10] Renaming these difference groups was necessary to avoid name conflicts. Other difference groups are being processed simultaneously with *w+SC-ls1* and *w+SC-ls2*.

■ CHAPTER 5
THE RECREL VALUES

5.1 INTRODUCTION

In this chapter we will examine the whole set of 61,075 values RECREL produces. We will start at a rather general level, by listing all distinct RECREL values and seeing how they are distributed: what is the number of SC pairs sharing a given value, what is the percent of all SC pairs having a value equal to or lower than that value, etc? After this, the value groups #n/#2-#12, $2 \leq n \leq 12$, will give us an approximate idea about what sort of RECREL values SCs of different cardinalities usually produce. Within each cardinality-class, we will identify the SCs having individual value groups X/#2-#12 with lowest and highest average values. The #n/#m value group information will follow, containing averages, minimum and maximum values, the SC pairs producing the extrema, and the numbers of distinct values in each value group. After this, RECREL will be compared with a modification of Lewin's REL measure.

In sections 5.3 - 5.7 we will investigate how RECREL similarity correlates with other aspects of SC similarity. Our starting point is a routine observation in the pcset-theoretical literature: pairs of SCs enjoying a given relation can be deemed similar. We collect all pairs of SCs enjoying the given relation, perform RECREL comparisons on them and draw conclusions. Typically, we observe whether the values are lower than average, whether they are consistently low or fluctuate between low and high values, whether SCs of small cardinalities produce results different from those of large cardinalities, whether the classes are each other's closest RECREL counterparts, etc. We will examine five types of case: inversionally related

SCs, Z-related SCs, complement pairs, SCs of cardinality n and their subset-classes of cardinality n-1, and M-related SCs. The results are summarized in section 5.8.

In the tables given in this chapter, average values, percentiles, etc., will usually be rounded to the nearest integer. At times, if greater accuracy is thought to be of benefit, decimal numbers will also be used. Many tables identify individual SC pairs with values of special interest (like minima or maxima). Such a value may be unique for the pair, or it may be shared by a group of pairs.

5.2 THE RECREL VALUE GROUP #2-#12/#2-#12

RECREL, like TMEMB, ATMEMB, REL and T%REL, is a total measure being able to discriminate between inversionally related and Z-related SCs. The scale of values comprises the integers between 0 and 100, inclusive. Some numbers between the extrema do not appear in the scale, however. The values from the 61,075 SC pair comparisons are distributed so that the number of distinct values is 89. They are listed in Table 5.1.

The average of all values is approximately 40. The median, having half of the values below it and the other half above it, is 36. As can be seen from Table 5.1, the number of SC pairs with the value 0, indicating maximal similarity, is 15. Among these are the seven pairs of inversionally related 3-pc classes, having always identical 2C%Vs, as well eight other pairs: {12-1,11-1}, {6-35,5-33}, {4-9,3-5A}, {4-9,3-5B}, {4-25,3-8A}, {4-25,3-8B}, {4-28,3-10} and {3-12,2-4}. Among other notably low values are RECREL(6-20,5-21A) = RECREL(6-20,5-21B) = 2, RECREL(4-2A,4-2B) = RECREL(4-22A,4-22B) = RECREL(5-21A,5-21B) = 4, and RECREL(8-28,7-31A) = RECREL(8-28,7-31B) = 6, etc.

Cases like these have an extremely small representation, since values up to 10 constitute only a fraction of one percent of all values. The values up to 20 still constitute less than 3% of all values. Above twenty the frequencies of the values start to grow rapidly, to the extent that values up to 50 constitute already more than 80% of all values. A full quarter of SC pairs have their values within the range 32-37 only. The value with the highest frequency is 32, belonging to 2,786 pairs. The most dissimilar 10% of the SC pairs is represented by a broad range of values from approximately 58 to 100. The value 100, indicating maximal dissimilarity and taking place only between SCs having completely disjoint subset-class contents, belongs to 220 pairs. Most of these pairs contain at least one dyad class.

TABLE 5.1: RECREL value distribution. The four columns list (1) the RECREL values, (2) the number of SC pairs having each value, (3) the percentual shares the column 2 entries have of the total number of 61,075 SC pairs, (4) the percentage of all 61,075 values which are equal to or lower than the corresponding values in column 1.

VAL	#	%	Pctl.	VAL	#	%	Pctl.	VAL	#	%	Pctl.
0	15	0.02	0.02	33	2478	4.06	37.28	63	218	0.36	92.99
2	2	0.00	0.03	34	2542	4.16	41.44	64	115	0.19	93.17
4	3	0.00	0.03	35	2755	4.51	45.95	65	120	0.2	93.37
6	2	0.00	0.04	36	2451	4.01	49.97	66	47	0.08	93.45
7	1	0.00	0.04	37	2477	4.06	54.02	67	686	1.12	94.57
8	15	0.02	0.06	38	2217	3.63	57.65	68	184	0.3	94.87
9	14	0.02	0.09	39	2069	3.39	61.04	69	38	0.06	94.93
10	16	0.03	0.11	40	1913	3.13	64.17	70	171	0.28	95.21
11	3	0.00	0.12	41	1703	2.79	66.96	71	137	0.22	95.44
12	20	0.03	0.15	42	1714	2.81	69.77	72	58	0.09	95.53
13	52	0.09	0.23	43	1272	2.08	71.85	73	238	0.39	95.92
14	73	0.12	0.35	44	1331	2.18	74.03	74	6	0.01	95.93
15	115	0.19	0.54	45	1196	1.96	75.99	75	71	0.12	96.05
16	161	0.26	0.81	46	1099	1.8	77.79	76	134	0.22	96.27
17	253	0.41	1.22	47	925	1.51	79.3	78	4	0.01	96.28
18	282	0.46	1.68	48	792	1.3	80.6	79	55	0.09	96.37
19	322	0.53	2.21	49	769	1.26	81.86	80	512	0.84	97.2
20	461	0.75	2.96	50	1265	2.07	83.93	81	230	0.38	97.58
21	501	0.82	3.78	51	533	0.87	84.8	82	179	0.29	97.87
22	683	1.12	4.9	52	500	0.82	85.62	83	297	0.49	98.36
23	877	1.44	6.34	53	449	0.74	86.35	86	199	0.33	98.69
24	952	1.56	7.9	54	554	0.91	87.26	87	219	0.36	99.04
25	1170	1.92	9.81	55	477	0.78	88.04	89	26	0.04	99.09
26	1467	2.4	12.21	56	422	0.69	88.73	90	223	0.37	99.45
27	1680	2.75	14.97	57	396	0.65	89.38	91	7	0.01	99.46
28	1728	2.83	17.79	58	370	0.61	89.99	92	14	0.02	99.49
29	1843	3.02	20.81	59	294	0.48	90.47	93	78	0.13	99.61
30	2358	3.86	24.67	60	622	1.02	91.49	95	16	0.03	99.64
31	2436	3.99	28.66	61	266	0.44	91.92	100	220	0.36	100.0
32	2786	4.56	33.22	62	431	0.71	92.63				

Table 5.2, listing the value groups #n/#2-#12, 2 ≤ n ≤ 12, shows that a set-class representing a small cardinality, when compared to the rest of the SC universe, produces on average considerably higher values than does a SC of a large cardinality. The three value groups #n/#2-#12, where n is 12, 11 or 10, produce a uniform average value, 33. This figure is only about 40% of the corresponding dyad class average 83. Also, among SCs of large cardinalities, all individual value groups X/#2-#12 produce reasonably similar averages. For example the average 33 of the value group 9-11B/#2-#12, being the lowest one for any nonad class, is only slightly lower than the highest nonad class average 35, being that of the value group 9-12/#2-#12.

Value groups #n/#2-#12, 3 ≤ n ≤ 6, contain much more variety. For example the lowest hexad class average, 34, which belongs to the value group 6-Z40B/#2-#12, is already considerably lower than the corresponding maximum, 59 of 6-35/#2-#12. The largest difference between lowest and highest averages is in the cardinality-class 4. The minimum, 37 of 4-Z29B/#2-#12, is only about half of the maximum, 73 of 4-28/#2-#12. The very highest individual average, 91, belongs to SC 2-6. Even with its closest RECREL counterparts the "tritone class" produces very high values, from 67 upwards. Along with 2-6, other transpositionally symmetric SCs, such as 3-12, 4-28, 6-35 and 8-28, are also well represented among the SCs with highest averages.[1]

TABLE 5.2: Aspects of the RECREL value groups #n/#2-#12 and the individual value groups X/#2-#12. The seven columns list (1) the value groups #n/#2-#12, (2) numbers of values in the value groups in column 1, (3) average values of the value groups in column 1, (4) the SCs X of cardinality n with the lowest individual value group X/#2-#12 averages, (5) the averages belonging to the SCs in column 4, (6) the SCs X of cardinality n with the highest individual value group X/#2-#12 averages, (7) the averages belonging to the SCs in column 6.

val.gr.	#	average	min-SC	ind.min	max-SC	ind.max
#12/#2-#12	349	33	12-1	33	12-1	33
#11/#2-#12	349	33	11-1	33	11-1	33
#10/#2-#12	2079	33	10-1	33	10-6	33
#9/#2-#12	6460	34	9-11B	33	9-12	35
#8/#2-#12	14104	35	8-Z29B	33	8-28	38
#7/#2-#12	20889	36	7-Z36B	33	7-33	42
#6/#2-#12	24760	38	6-Z40B	34	6-35	59
#5/#2-#12	20889	40	5-Z38B	36	5-33	58
#4/#2-#12	14104	46	4-Z29B	37	4-28	73
#3/#2-#12	6460	56	3-11B	49	3-12	82
#2/#2-#12	2079	83	2-5	82	2-6	91

We criticized the two preliminary versions of RECREL, $\%REL_n$ and T%REL, for producing values which were too low and too high, respectively (sections 3.4.2.1 and 3.7.4.1). When we compare the $\%REL_2$ and T%REL value group information tables 3.2 and 3.18 to the corresponding RECREL Table 5.3, we note that the RECREL averages are usually somewhere between their $\%REL_2$ and T%REL counterparts. For example the $\%REL_2$ averages in the value groups #4/#n, $5 \leq n \leq 9$, vary approximately between 32 and 39, their T%REL counterparts approximately between 67 and 70. The corresponding RECREL averages lie between 43 and 46. The $\%REL_2$ averages in the value groups #7/#n, $7 \leq n \leq 9$, lie approximately between 11 and 13, their T%REL counterparts approximately between 48 and 53. The corresponding RECREL averages are between 27 and 28. And so on.

Predictably, RECREL repeats the general observation of many similarity measures, by deeming SCs of the smallest cardinalities highly dissimilar on average, and SCs of the largest cardinalities highly similar on average.

---

[1] For details on transpositionally symmetric SCs, see for example Rahn (1980:92) or Castrén (1989:55-87).

TABLE 5.3: The RECREL value groups. Top row of a cell: minimum and maximum values. Middle row: SC pairs with the minimum and maximum values, respectively. Ordinal numbers, hyphens and Zs omitted from the SC names. If n ≠ m in #n/#m, the SC to the left of the virgule is of larger cardi - nality. Bottom row: the average value and the number of distinct values. Cells #11/#11 and #12/#12 omitted as the comparison groups contain only pairs {11-1,11-1} and {12-1,12-1}, respectively. C = 100.

	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11
#2	C C 1/2 1/2 C 1									
#3	0 C 12/4 1/3 83 4	0 C 2A/2B 1/10 61 4								
#4	33 C 28/3 1/4 83 5	0 C 9/5A 1/12 56 13	4 83 2A/2B 1/28 46 33							
#5	60 C 1/1 1/5 83 5	15 90 33/8A 1/12 56 29	10 80 33/21 1/28 46 61	4 65 21A/21B 7A/33 37 54						
#6	60 C 20/4 1/6 83 7	15 87 35/8A 1/12 55 33	10 80 35/21 1/28 45 56	0 66 35/33 35/7A 38 55	8 64 14A/14B 5A/35 33 52					
#7	71 95 1/1 1/6 83 6	33 86 1/2A 1/12 55 30	27 76 1/2A 1/28 44 42	21 61 1/2A 4A/33 37 35	13 62 31A/27A 7A/35 34 42	9 43 7A/7B 31A/33 27 34				
#8	71 93 28/3 1/6 83 7	39 86 1/2A 1/12 55 29	29 75 28/12A 1/28 43 38	24 61 28/10A 6/33 37 30	15 62 28/13 9/35 33 37	6 45 28/31A 28/33 28 32	11 38 19A/19B 21/28 23 28			
#9	75 92 12/4 1/6 83 6	44 83 3A/3A 1/12 55 20	30 75 5A/15A 1/28 43 34	30 59 1/12 1/33 37 22	25 59 1/3A 1/35 32 23	19 39 12/13A 5A/33 27 20	10 37 12/19A 1/28 23 23	8 27 8A/8B 10/12 17 15		
#10	80 91 1/1 1/6 83 4	47 82 1/2A 1/12 55 11	31 73 6/15A 1/28 43 29	31 59 1/11A 1/33 36 20	27 59 1/10A 1/35 32 21	23 38 1/12 1/33 27 14	17 35 6/15A 1/28 22 16	12 26 6/5A 1/12 17 12	9 13 1/5 1/6 11 3	
#11	82 91 1/1 1/6 84 2	48 82 1/2A 1/12 55 6	32 73 1/15A 1/28 43 13	31 58 1/11A 1/33 36 12	27 58 1/11A 1/35 32 14	23 37 1/12 1/33 27 10	19 34 1/15A 1/28 22 10	15 24 1/2A 1/12 16 4	8 13 1/1 1/6 9 3	
#12	82 91 1/1 1/6 84 2	48 82 1/2A 1/12 55 6	32 73 1/15A 1/28 43 13	31 58 1/11A 1/33 36 12	27 58 1/11A 1/35 32 14	23 37 1/12 1/33 27 10	19 34 1/15A 1/28 22 10	15 24 1/2A 1/12 16 4	8 13 1/1 1/6 9 3	0 0 1/1 1/1 0 1

In RECREL value groups #n/#m having a large difference between n and m, neither the averages nor the minimum values are especially low. The problem of %REL_2, i.e., counterintuitively low values between SCs of greatly differing cardinalities, does not arise here. For example the lowest value in any of the value groups #2/#n, $5 \leq n \leq 12$, is 60. Among the value groups #3/#n, $7 \leq n \leq 12$, the lowest value is 33, among the value groups #4/#n, $8 \leq n \leq 12$, the minimum is 29, etc.

By examining the columns of Table 5.3 from top to bottom, we see that the average values in the highest value groups (#n/#n) are usually of the same approximate size as those in the lower value groups (#n/#n+1, #n/#n+2,... #n/#12). In contrast, the minimum and maximum values in the two highest groups tend to be clearly lower and slightly higher respectively than their counterparts in the lowest groups of the corresponding column. This indicates that the ranges of values are largest when we compare exactly or approximately equal-sized SCs, which is an intuitively acceptable result.

The transpositionally symmetric SCs often produce a small handful of exceptionally low values and a great number of very high values with other SCs. As a result, they have a disproportionately large representation among the pairs with minimum and maximum values in Table 5.3. Consider, for example, column #6. Every SC pair with the maximum value in all seven cells contains SC 6-35, the whole-tone class. Moreover, in three cells out of five in the row #6, 6-35 participates in producing the *lowest* value in the value group. In the cells of row #8, the octatonal class 8-28 participates five times in producing the pair with the minimum value, and in column #8 the same class is found in every pair having the maximum value. Its complement, the diminished seventh class 4-28, is in every maximum pair of column #4, and the complement classes 3-12 and 9-12 are prevalent in the maximum pairs of columns #3 and #9, respectively, etc.

5.2.1 RECREL and REL'

Let use briefly compare RECREL and REL. The latter was the only previously presented measure we approved during the evaluations in chapter 3. Because original REL values are somewhat difficult to relate to RECREL values - the scale is from 0 to 1 and increasing values indicate increasing similarity - we modify REL by subtracting each value from 1 and multiplying the difference by 100. We will call the modification REL'. Its scale of values is identical to that of RECREL, the end points being 0 and 100 and increasing values indicating increasing dissimilarity. The results are compiled under T_n-classification. The value group information is given in Table 5.4.

TABLE 5.4: The REL' value groups #n/#m, 3 ≤ n,m ≤ 9. Lower values indicate higher degrees of simi - larity. In each table cell, clockwise from top left: the lowest and highest values, the number of distinct values, the average. C = 100.

	#3	#4	#5	#6	#7	#8	#9
#3	25 C 68.09 7						
#4	1 C 58.37 21	15 87 54.18 36					
#5	20 84 55.18 31	8 88 49.82 63	14 81 45.75 57				
#6	23 85 55.37 38	12 90 50.36 57	2 83 44.14 68	14 86 40.75 57			
#7	39 80 55.93 30	35 77 51.7 38	27 71 44.82 43	13 75 38.25 57	9 55 32.47 38		
#8	48 81 57.23 22	45 77 54.44 28	35 74 48.16 34	18 78 40.46 47	5 55 30.72 44	10 50 25.15 36	
#9	53 76 58.6 14	51 79 57.09 21	46 69 51.91 19	36 73 44.27 29	22 51 32.63 29	6 55 22.06 33	6 30 14.19 16

The two tables show that in a clear majority of cases, RECREL averages, minima and maxima are lower than their REL' counterparts. This, no doubt, is due to applying the non-common subset-class criterion C6. REL' value groups have lower averages in only two cases, #8/#9 and #9/#9. RECREL maxima are equal to or above their REL' counterparts only in the seven value groups #3/#n. In the large-cardinality value groups REL' maxima are almost suspiciously high. In the group #8/#9, for ex - ample, the maximum is 55.

An interesting observation is that the ways in which the corresponding entries relate to each other do not always seem to create patterns. This makes comparisons between the measures extremely complicated. In the two columns #4/#n, for example, RECREL minima are usually clearly lower than their REL' counterparts. The REL' value group #4/#5 minimum, however, is below its counterpart, and the #4/#6 entry for RECREL is only slightly lower. In the two rows #9/#n, in turn, the differences between corresponding averages are small when n is 3, 8 and 9. When n is 4, 5 or 6, the differences are well over 10, etc. It is difficult, on the basis of the value group analysis, to regard one measure as better than the other.

5.3 RECREL VALUES BETWEEN INVERSIONALLY RELATED SCS

RECREL values belonging to pairs of inversionally related SCs are usually not uniform in a single value group.[2] This was also the case with ATMEMB, REL and

[2] Inversionally symmetric SCs, reproducing themselves under inversion, are excluded from the com-

T%REL. Table 5.5 shows how the RECREL values belonging to inversionally related pairs vary within the value groups #n/#n, 3 ≤ n ≤ 9. #3/#3 is the only value group within which the values produced by the I-pairs indeed are uniform, being instances of the value 0. In the other value groups, the distances between the lowest and highest values range between an extremely narrow 1, in #9/#9, and a rather wide 21, in #4/#4. In the latter case, the two pairs with the maximum value 25 contain the I-related "all-interval tetrachords" {4-Z29A,4-Z29B} and {4-Z15A,4-Z15B}. 25 is the highest value among all 128 I-pairs.

Comparing the average values (column 7) to their counterparts in the corresponding #n/#n cells in Table 5.3 shows that the I-pair averages are considerably below the #n/#n averages. For example the highest I-pair average within a single value group, 17 in #6/#6, is only about half of the entire #6/#6 average, 33. Even the I-pair maxima in column 6 of Table 5.5 are clearly below the #n/#n averages. Note also that in Table 5.3, every comparison group #n/#n containing inversionally related pairs has one such pair as the pair with the lowest value. The percentile column 8 in Table 5.5 indicates that within its comparison group, a pair with the average I-pair value or lower belongs to a very small percentage of most similar pairs. According to Table 5.1, a pair with a value equal to or lower than the average of all 128 I-pair values, 14, belongs to 0.35% of most similar pairs.

TABLE 5.5: RECREL values indicating the highest, lowest and average degrees of similarity among pairs of inversionally related SCs. The eight columns list (1) the comparison groups #n/#n containing the inversionally related SCs and, as the lowest entry, the cardinality-class range #3-#9, (2) the numbers of inversionally related SC pairs in the comparison groups in column 1, (3) the most similar inversionally related SC pairs in the comparison groups, (4) the values belonging to the pairs in column 3, (5) the most dissimilar inversionally related SC pairs in the comparison groups, (6) the values belonging to the pairs in column 5, (7) the average values of all inversionally related SCs in the comparison groups, (8) the percentages of all SC pairs in the corresponding comparison groups with values equal to or lower than the values in column 7.

C.Group	#	Most Simil.	Value	Most Dissimil.	Value	Avg.	Pctl.
#9/#9:	7	{9-8A,9-8B}	8	{9-11A,9-11B}	9	9	4
#8/#8:	14	{8-19A,8-19B}	11	{8-Z29A,8-Z29B}	18	14	1
#7/#7:	28	{7-7A,7-7B}	9	{7-Z38A,7-Z38B}	17	14	1
#6/#6:	30	{6-14A,6-14B}	8	{6-Z40A,6-Z40B}	22	17	2
#5/#5:	28	{5-21A,5-21B}	4	{5-29A,5-29B}	20	14	1
#4/#4:	14	{4-2A,4-2B}	4	{4-Z29A,4-Z29B}	25	14	2
#3/#3:	7	{3-2A,3-2B}	0	{3-11A,3-11B}	0	0	4
#3-#9:	128	{3-2A,3-2B}	0	{4-Z29A,4-Z29B}	25	14	-

Given, in turn, each inversionally non-symmetric SC X, its inversionally related class I(X) and its individual RECREL value group X/#2-#12, the average number of val -

parisons in the present section.

ues below the RECREL(X,I(X)) value in the value group is only 0.8. This figure is considerably lower than its counterparts produced by ATMEMB, REL and T%REL, being 11, 18 and 4, respectively.[3] It is always true that if X/#2-#12 contains n values below the RECREL(X,I(X)) value, also I(X)/#2-#12 will contain n values below the RECREL(I(X),X) value.

Out of the 128 I-related pairs, no less than 102 are those where the SCs are each other's closest RECREL counterparts. For the other three similarity measures mentioned above, the corresponding numbers are 29, 11 and 62 out of 128, respectively. There are only 26 I-pairs where the SCs are not each other's closest RECREL counterparts. In value group 6-27A/#2-#12 there are 12 values below the RECREL(6-27A,6-27B) value 21, the largest such number. Among pairs containing 6-27A and with a value below 21 are, for example, {6-27A,7-31A}, with value 13, and {6-27A,8-28}, with value 17. Other SCs which have many values below the I-counterpart value are, for example, 6-5A and 6-18A. Both have 10 values below the I-counterpart value.

Given an inversionally symmetric SC S and two inversionally non-symmetric SCs X and Y, it is always true that RECREL(S,X) = RECREL(S,I(X)), and that RECREL(X,Y) = RECREL(I(X),I(Y)). RECREL(X,Y) may or may not be RECREL(X,I(Y)).[4]

5.4 RECREL VALUES BETWEEN Z-RELATED SCS

When Z-related SCs are investigated under T_n-classification, the task differs from the one under T_n/I-classification in one respect, i.e., one SC can have two Z-counterparts. Thus, for example, one of the "all-interval tetrachords," 4-Z15A, is Z-related to both 4-Z29A and 4-Z29B. An inversionally non-symmetric SC with Z-counterparts usually has two of them, the counterparts constituting an I-pair. An inversionally symmetric SC with Z-counterparts has usually just one, also an inversionally symmetric SC. There are two exceptions, however. The SC 5-Z12, being inversionally symmetric, is Z-related to the inversionally related classes 5-Z36A and 5-Z36B. Likewise, the inversionally symmetric SC 7-Z12 is Z-related to the inversionally related classes 7-Z36A and 7-Z36B. Altogether, the number of Z-related pairs (Z-pairs) under T_n-classification is 61. The number of distinct SCs participating in these pairs is 72.

[3] For ATMEMB and REL we counted values *exceeding* the {X,I(X)} value, of course.

[4] The formalisations were obtained through exhaustive computer searches.

The RECREL values belonging to Z-related pairs in a single #n/#n comparison group are not uniform. This was also the case with ATMEMB, REL and T%REL. Obviously, the possibility of having two Z-counterparts gives rise to a question: Given an inversionally non-symmetric SC X and its two inversionally related Z-counterparts Y_1 and Y_2, is RECREL(X,Y_1) always RECREL(X,Y_2)?[5] The answer is no. The two values are different in every case. Let the value RECREL(X,Y_1) be the higher one of the two. On average, RECREL(X,Y_1) - RECREL(X,Y_2) = 6.

The smallest difference is produced by the case RECREL(6-Z44A,6-Z19A) - RECREL(6-Z44A,6-Z19B) = 15 - 14 = 1. The highest difference, in turn, is produced by RECREL(4-Z15A,4-Z29A) - RECREL(4-Z15A,4-Z29B) = 25 - 8 = 17. We saw in section 5.3 above that the two inversionally related "all-interval tetrachord" pairs had the highest value, 25, among all I-pairs. Each of these four SCs is considerably closer to one of its two Z-counterparts than to its I-counterpart.

Comparing corresponding column 7 entries in tables 5.5 and 5.6 shows that the Z-pairs do not have quite as low average values as did the I-related pairs. Still, pairs with the average Z-pair values or below are among the most similar in their comparison groups (Table 5.6, column 8). The highest average, 22, belongs to the 6-pc Z-pairs, but even that value places a pair in the 10% of closest pairs in the comparison group #6/#6.

TABLE 5.6: RECREL values indicating the highest, lowest and average degrees of similarity among pairs of Z-related SCs. The eight columns list (1) the comparison groups #n/#n containing the Z-related SCs and, as the lowest entry, the cardinality-class range #4-#8, (2) the numbers of Z-related SC pairs in the comparison groups in column 1, (3) the most similar Z-related SC pairs in the comparison groups, (4) the values belonging to the pairs in column 3, (5) the most dissimilar Z-related SC pairs in the comparison groups, (6) the values belonging to the pairs in column 5, (7) the average values of all Z-related SC pairs in the comparison groups, (8) the percentages of all SC pairs in the corresponding comparison groups with values equal to or lower than the values in column 7.

C.Group	#	Most Simil.	Value	Most Dissimil.	Value	Avg.	Pctl.
#8/#8	4	{8-Z29A,8-Z15B}	12	{8-Z29A,8-Z15A}	18	15	5
#7/#7	7	{7-Z37,7-Z17}	13	{7-Z38A,7-Z18B}	21	17	4
#6/#6	39	{6-Z38,6-Z6}	13	{6-Z46B,6-Z24A}	27	22	10
#5/#5	7	{5-Z37,5-Z17}	12	{5-Z38B,5-Z18A}	25	19	3
#4/#4	4	{4-Z29B,4-Z15A}	8	{4-Z29B,4-Z15B}	25	16	2
#4-#8	61	{4-Z29B,4-Z15A}	8	{6-Z46B,6-Z24A}	27	20	-

The values of all 61 Z-pairs range between RECREL(4-Z29B,4-Z15A) = 8 and RECREL(6-Z46B,6-Z24A) = 27. Within the value groups #n/#n, the distances between the lowest (column 4) and highest (column 6) Z-pair values vary from 6

[5] Cases where X is inversionally symmetric do not have to be examined, because RECREL(X,Y_1) is automatically RECREL(X,Y_2).

(#8/#8) to 17 (#4/#4). Even the maxima are comfortably below the value group #n/#n averages in Table 5.3. In both column 3 and column 5, the SCs in the pairs belonging to #8/#8 and #7/#7 are complements of the classes in the pairs belonging to #4/#4 and #5/#5, respectively.

Given, in turn, each Z-related SC pair $\{Z_1,Z_2\}$ and the two individual RECREL value groups Z_1/#2-#12 and Z_2/#2-#12, the average number of values below the RECREL(Z_1,Z_2) value in each of the two value groups is 10. The corresponding number for T%REL was the same, for ATMEMB and REL (number of values *exceeding* the $\{Z_1,Z_2\}$ value) 20 and 30, respectively.

There are 32 SCs which produce their lowest RECREL values with their Z-counterparts, or if they have two, with the closer counterpart. SC 6-Z10A is at the other end of the scale, as it was for ATMEMB and REL, too. In the value group 6-Z10A/#2-#12 there are 49 values below the RECREL(6-Z10A,6-Z39A) value 27. And as with ATMEMB, REL and T%REL, the numbers are not necessarily the same for the two Z-counterparts. In 6-Z39A/#2-#12 there are 40 values below the RECREL(6-Z39A,6-Z10A) value.

5.5 RECREL VALUES BETWEEN COMPLEMENT CLASSES

Under T_n-classification, each inversionally non-symmetric SC X has both a complement class X_C and an inverted complement class (*I-complement*) I(X_C) so that $X_C \neq$ I(X_C). Usually the complement class of an A-type class is a B-type class, but in 14 complement pairs both SCs are of either A-type or B-type. Thus, for example, the complement class of 5-11A is 7-11A, that of 5-11B 7-11B.

TABLE 5.7: RECREL values indicating the highest, lowest and average degrees of similarity among pairs of complement classes. The eight columns list (1) the comparison groups #n/#(12-n) containing the complement pairs and, as the lowest entry, the cardinality-class range #2-#10, (2) the numbers of complement pairs in the comparison groups in column 1, (3) the most similar complement pairs in the comparison groups, (4) the values belonging to the pairs in column 3, (5) the most dissimilar complement pairs in the comparison groups, (6) the values belonging to the pairs in column 5, (7) the average values of all complement pairs in the comparison groups, (8) the percentages of all SC pairs in the corresponding comparison groups with values equal to or lower than the values in column 7. Excluded from the table are the self-complementing hexad classes, producing trivially value 0, and the complement pairs {12-1,0-1} and {11-1,1-1}, which do not have RECREL values.

C.Group	#	Most Simil.	Value	Most Dissimil.	Value	Avg.	Pctl.
#2/#10	6	{2-5,10-5}	80	{2-6,10-6}	89	82	83
#3/#9	19	{3-11B,9-11A}	44	{3-12,9-12}	75	51	50
#4/#8	43	{4-Z29B,8-Z29A}	33	{4-28,8-28}	57	37	28
#5/#7	66	{5-21B,7-21A}	24	{5-28B,7-28B}	33	29	8
#6/#6	36	{6-Z38,6-Z6}	13	{6-Z46B,6-Z24A}	27	22	10
#2-#10	170	{6-Z38,6-Z6}	13	{2-6,10-6}	89	34	-

Two categories of complement pairs do not contribute to Table 5.7. First, the self-complementing hexachord classes, producing trivially the value 0 with themselves; second, the pairs {12-1,0-1} and {11-1,1-1}, which do not have RECREL values at all.

Among 170 complement pairs the values are between 13 and 89, the average being 34. Columns 3 and 4 show, predictably, that the very lowest complement value belongs to a hexad class pair. When the differences between the cardinalities are large, even the minima are high.

The complement pair averages in column 7, belonging to the five value groups #n/#(12-n), are all below the corresponding averages in Table 5.3. So are the #5/#7 and #6/#6 maxima in column 6. The #4/#8, #3/#9 and #2/#10 maxima, in contrast, are above the Table 5.3 averages, the first two even considerably. The distances between the lowest and highest values vary from 9 (in value groups #2/#10 and #5/#7) to 31 (in value group #3/#9).

The percentiles (column 8) fluctuate strongly. For example a pair with the average #5/#7 complement pair value, 29, belongs to 8% of closest pairs in the entire comparison group, whereas the #3/#9 value, 51, is exactly the median. The transpositionally symmetric SCs are again represented among the most dissimilar pairs (column 5).

In section 5.4 above, we examined how many values below the RECREL(Z_1,Z_2) value each pair of Z-related SCs {Z_1,Z_2} had in their two individual value groups Z_1/#2-#12 and Z_2/#2-#12. Here, since the differences between the cardinalities of complement classes {X,X_C} vary, pairs with large differences tend to have high values and, possibly, individual value groups X/#2-#12 and X_C /#2-#12 with a large number of values below RECREL(X,X_C). To remove the distortion this tendency might cause in the results, we limit the individual value groups. If X is of cardinality n and X_C of cardinality (12-n), the two individual value groups to be studied will be X/#(12-n) and X_C/#n, respectively.

Thus, for example, given the complement pair {3-1,9-1}, we observe the RECREL value, 59, and count how many values below it we have in the two value groups 3-1/#9 and 9-1/#3. The results show whether the complement-related SCs are each other's closest RECREL counterparts in each other's cardinality-classes, a reasonably uniform test for all complement pairs.

Given in turn each pair of complement classes {X,X_C}, $X \neq X_C$, and the two individual RECREL value groups X/#(12-n) and X_C /#n where n and (12-n) are the cardinalities of X and X_C, respectively, the average number of values below the RECREL(X,X_C) value in each of the two value groups is six.

There are 74 SCs for which the lowest RECREL counterpart in the cardinality-class of the complement is the complement itself. 74 is only 21% of all SCs between cardinalities 2 and 10. At the other extreme, in the individual RECREL value group 8-28/#4, the number of values below the RECREL(8-28,4-28) value, 57, is 36. As the total number of tetrad classes is 43, more than 80% of them are more similar to 8-28 than its complement is. Furthermore, in the individual value group 9-12/#3, the number of values below the RECREL(9-12,3-12) value, 75, is 17. Astonishingly, along with 3-10, 3-12 is the most dissimilar triad class counterpart for 9-12.

From the point of view of the smaller SCs the situation is quite different. In the value groups 4-28/#8 and 3-12/#9, the values with the complement classes are the lowest ones. This sort of dramatic asymmetry in the results is connected to subset-class instance distribution. SCs of small cardinalities with highly uneven instance distributions, such as 3-12 and 4-28, produce high RECREL values with all SCs of large cardinalities, as the latter ones tend to have even dyad and tetrad class instance distributions. In contrast, SCs of small cardinalities which have the most even instance distributions possible, produce with their complements values that are usually at least reasonably close to the lowest ones in the value groups X_C/#n. For example RECREL(3-11B,9-11A) = 44. This is the lowest value in the two value groups 3-11B/#9 and 9-11A/#3. There are cases, however, where the asymmetry between the two sets of values below the $\{X,X_C\}$ value is of the reversed kind. For example RECREL(5-Z38A,7-Z38B) = 32. The value group 5-Z38A/#7 contains 28 values below 32. In the value group 7-Z38B/#5, in contrast, the number of values below 32 is only 9.

Let us also briefly investigate RECREL values between SCs and their inverted complements. There are three types of I-complement pairs not contributing to the table. See Table 5.8 legend. The interesting aspect about the 158 I-complement pairs is revealed by comparing tables 5.7 and 5.8.[6] Entries in the latter table are either similar or lower than their counterparts in the former. Generally, the I-complement pairs represent slightly closer RECREL similarity than the complement pairs. We have no explanation for this observation.

According to RECREL, complement pairs represent anything but a consistent degree of similarity. This observation has interesting implications, as well-known theoretical concepts and analytical principles emphasize the importance of the complement relation. Forte, deeming the complement of a SC a "reduced or enlarged replica" of that SC (1973a:78), gives the complement relation an important role in his set-complexes (Ibid., 93-7), and suggests that if a SC is to be structurally significant in an atonal piece of Schönberg, also its complement should occur throughout the piece (Forte 1972:45 and 1982:132-5).

---

[6] Note that the set of all complement pairs and the set of all I-complement pairs intersect. Given an inversionally symmetric SC S, $S_C = I(S_C)$.

TABLE 5.8: RECREL values indicating the highest, lowest and average degrees of similarity among pairs of inverted complement classes. The eight columns list (1) the comparison groups #n/#(12-n) containing the inverted complement pairs and, as the lowest entry, the cardinality-class range #2-#10, (2) the numbers of inverted complement pairs in the comparison groups in column 1, (3) the most similar inverted complement pairs in the comparison groups, (4) the values belonging to the pairs in column 3, (5) the most dissimilar inverted complement pairs in the comparison groups, (6) the values belonging to the pairs in column 5, (7) the average values of all inverted complement pairs in the comparison groups, (8) the percentages of all SC pairs in the corresponding comparison groups with values equal to or lower than the values in column 7. Three types of case do not contribute to the table: (a) the self-complementing, inversionally symmetric hexad classes (being their own inverted complements), (b) inversionally non-symmetric hexad classes whose complements are their inversionally related classes, (also their own inverted complements), (c) the complement pairs {12-1,0-1} and {11-1,1-1}, not having RECREL values.

C.Group	#	Most Simil.	Value	Most Dissimil.	Value	Avg.	Pctl.
#2/#10	6	{2-5,10-5}	80	{2-6,10-6}	89	82	83
#3/#9	19	{3-11B,9-11B}	44	{3-12,9-12}	75	51	50
#4/#8	43	{4-Z29B,8-Z29B}	30	{4-28,8-28}	57	36	20
#5/#7	66	{5-23B,7-23B}	24	{5-Z12,7-Z12}	32	26	3
#6/#6	24	{6-14B,6-14A}	8	{6-Z48,6-Z26}	24	20	4
#2-#10	158	{6-14B,6-14A}	8	{2-6,10-6}	89	33	-

5.6 RECREL VALUES BETWEEN SCS OF CARDINALITY N AND THEIR SUBSET-CLASSES OF CARDINALITY N-1

RECREL, being based on the extent of mutual embedding of subset-classes in two SCs, is closely connected to the inclusion relation. Given two SCs X and Y of cardinalities n and m so that Y is included in X, we may then ask, is the fact that every subset-class in Y is also a subset-class in X reflected in the value RECREL(X,Y), when compared to all values in the value group #n/#m? Or, generally, do inclusion-related pairs have lower than average RECREL values within their value groups? In terms of differences between cardinalities, there are a number of "difference categories" we can study among the comparison groups: #n/#n-1, #n/#n-2, etc. As examining all inclusion relations is beyond our present scope, and as the inclusion relation becomes extremely common when the cardinality difference is sufficiently great - all dyad classes are subset-classes of all classes of cardinality 7 or higher, for example - we will concentrate on just one "difference category," the comparison groups #n/#n-1 and the inclusion-related SC pairs within them.

By the term *n-1 classes* let us refer to all SCs representing a cardinality one less than the cardinality n of some referential class. Accordingly, we will speak of *n-1 subset-classes*, *n-1 subset class averages* (average RECREL values between SCs of cardinality n and their subset-classes of cardinality n-1), *non-included n-1 classes*, etc. The triad classes are the SCs of smallest n whose n-1 classes we will examine. The upper limit for n is 10, as there is only one SC in each of the cardinality-classes 11 and 12.

The n-1 subset-class information is given in Table 5.9. The first observation we make is the same we already made in connection with the inversionally related and Z-related pairs and the complement pairs: the degree of similarity is not uniform within one value group. Some SCs of cardinality n can be considerably more similar to their n-1 subset-classes than some others.

In a few cases SCs are maximally similar to their n-1 subset-classes. For example the sole n-1 subset-class of SC 6-35 is 5-33. RECREL(6-35,5-33) = 0. SC 4-9 contains two instances of 3-5A and two of 3-5B. RECREL(4-9,3-5A) = RECREL(4-9,3-5B) = 0. Out of the fifteen SC pairs with RECREL value 0, eight are inclusion-related and belong to one of the comparison groups #n/#n-1. (Section 5.2).

TABLE 5.9: RECREL values indicating the highest, lowest and average degrees of similarity between SCs of cardinality n and their subset-classes of cardinality n-1. The nine columns list (1) the comparison groups #n/#n-1 containing the pairs and, as the lowest entry, the cardinality-class range #3-#10, (2) the numbers of SCs in the cardinality-classes n in column 1, (3) the SCs of cardinality n having the lowest average values with their subset-classes of cardinality n-1, (4) the average values belonging to the SCs in column 3, (5) the SCs of cardinality n having the highest average values with their subset-classes of cardinality n-1, (6) the average values belonging to the SCs in column 5, (7) the averages of average values between each SC X of cardinality n and all SCs of cardinality n-1 included in X, (8) the percentages of all SC pairs in the corresponding comparison groups with values equal to or lower than the values in column 7, (9) the averages of average values between each SC X of cardinality n and all SCs of cardinality n-1 not included in X.

C.Group	#	Most Sim.	Val	Most Dissim.	Val	Avg.	Pctl.	Avg.Non-Inc.
#10/#9	6	10-6	13	10-4	15	14	33	19
#9/#8	19	9-12	11	9-8B	20	18	13	25
#8/#7	43	8-28	6	8-Z29B	23	20	6	29
#7/#6	66	7-31B	15	7-28B	26	23	6	35
#6/#5	80	6-35	0	6-Z45	31	24	4	40
#5/#4	66	5-33	12	5-Z36B	39	29	8	48
#4/#3	43	4-28	0	4-Z29B	50	34	9	62
#3/#2	19	3-12	0	3-11B	67	59	4	100
#3-#10	342	6-35	0	3-11B	67	27	-	44

The distances between the lowest values (column 4) and the highest values (column 6) vary from a very narrow 2 (value group #10/#9) to an extremely broad 67 (value group #3/#2). All SCs in column 3 are either transpositionally symmetric or multiple-instance subset-classes of such: 7-31B of 8-28, 5-33 of 6-35. The SCs in column 5, by contrast, are classes with more even subset-class distributions. The averages (column 7), despite reaching as high a value as 59, are lower than their counterparts in the #n/#n-1 cells in Table 5.3, and often considerably so. The percentiles (column 8) show that in some comparison groups, the pairs with average n-1 subset-class values or lower belong to quite small groups of most similar pairs.

Comparisons between corresponding entries in columns 7 and 9 give an additional viewpoint. The average inclusion-related #n/#n-1 pair has a clearly lower value than the average #n/#n-1 pair without the inclusion relation. Not even the column 6 maxima reach their column 9 counterparts.

There are 342 SCs between the cardinalities 3 and 10. Out of these, 311 are such that all their closest n-1 RECREL counterparts are also their subset-classes. In 28 cases the lowest n-1 value is shared by a subset-class and a non-included class. There are only three cases where non-included n-1 classes are the closest n-1 counterparts alone. These three SCs, 5-Z17, 5-22 and 5-Z37, produce the value 28 with their closest included tetrad classes. Each produce a slightly lower value, 27, with two non-included tetrad classes. In the case of 5-Z17, for example, these classes are 4-7 and 4-17.

5.7 RECREL VALUES BETWEEN M-RELATED SCS

When the M operation and/or M-related SCs are being discussed from the point of view of our present interests, aspects of abstract structural similarity seem to prevail over those of intuitively experienced resemblance.[7] For example, writers point to predictable regularities in the ic contents of M-related SCs, or note that M, M_1, M_7 and M_{11} are the only operators on 12 pcs that are isomorphisms in the group-theoretical sense (Morris 1987:79-80, Rahn 1980:53-5). Doubts are even expressed about the possibility of deriving aurally similar pitch combinations from at least some M-related SCs (Morris 1987:79).

Under T_n-classification, the M-counterpart of a given SC X can be X itself, its inversionally related SC I(X), or some other SC Y. The pairs M(X) = X are excluded from the calculations below. In the following, when we refer to the M-related pairs, we mean the 134 SC pairs where $M(X) \neq X$.

The information in Table 5.10 shows that from the point of view of RECREL, the degrees of similarity among M-pairs in a single comparison group #n/#n can fluctuate significantly. In groups with more than one M-pair (all except #10/#10 and #2/#2), the distances between the lowest values (column 4) and the highest values (column 6) vary from 11 (#9/#9) to as high as 67 (#3/#3). The SCs in the most similar pairs (column 3) are often inversionally related to each other. The averages (column 7) are below their RECREL value group Table 5.3 counterparts, but in the

7 Following the terminological convention used in Morris (1987), by M we mean only what is in some sources identified as M_5. SC pairs related by other multiplication operations, such as M_7 and M_{11}, will not be examined in this section. For a description of the concept and related discussion, see for example Rahn (1980:53-6) and Morris (1987:65-6).

cases of the comparison groups #n/#n, $3 \leq n \leq 9$, the column 6 maxima can even be considerably higher than the averages in Table 5.3. Notably, all of the most dissimilar pairs (column 5) belong to the most often discussed category of M-related pairs, as each contains a "chromatic segment" and a "circle of fourths segment" of the same length.

TABLE 5.10: RECREL values indicating the highest, lowest and average degrees of similarity among pairs of M-related SCs. The SCs which are their own M-counterparts do not contribute to the table. The eight columns list (1) the comparison groups #n/#n containing the M-related SC pairs and, as the lowest entry, the cardinality-class range #2-#10, (2) the numbers of M-related SC pairs in the comparison groups in column 1, (3) the most similar M-related pairs in the comparison groups, (4) the values belonging to the pairs in column 3, (5) the most dissimilar M-related pairs in the comparison groups, (6) the values belonging to the pairs in column 5, (7) the average values of all M-related SC pairs in the comparison groups, (8) the percentages of all SC pairs in the corresponding comparison groups with values equal to or lower than the values in column 7.

C.Group	#	Most Simil.	Value	Most Dissimil.	Value	Avg.	Pctl.
#10/#10	1	{10-1,10-5}	9	{10-1,10-5}	9	9	7
#9/#9	7	{9-8A,9-8B}	8	{9-1,9-9}	19	13	13
#8/#8	15	{8-7,8-20}	13	{8-1,8-23}	27	18	21
#7/#7	28	{7-7A,7-7B}	9	{7-1,7-35}	37	20	10
#6/#6	32	{6-Z13,6-Z50}	13	{6-1,6-32}	41	22	10
#5/#5	28	{5-7A,5-7B}	7	{5-1,5-35}	53	25	10
#4/#4	15	{4-18A,4-18B}	14	{4-1,4-23}	54	30	21
#3/#3	7	{3-8A,3-8B}	0	{3-1,3-9}	67	28	4
#2/#2	1	{2-1,2-5}	100	{2-1,2-5}	100	100	100
#2-#10	134	{3-8A,3-8B}	0	{2-1,2-5}	100	23	-

Given in its turn each M-related SC pair $\{M_1,M_2\}$, $M_1 \neq M_2$, and the two individual RECREL value groups M_1/#2-#12 and M_2/#2-#12, the average number of values below the RECREL(M_1,M_2) value in each of the two value groups is 27. In every case, if M_1/#2-#12 contains n values below the RECREL(M_1,M_2) value, also M_2/#2-#12 will contain n values below the RECREL(M_2,M_1) value.

There are 32 pairs of M-related SCs such that the classes are each other's closest RECREL counterparts. At the other extreme is for example the pair {5-1,5-35}. Each of the value groups 5-1/#2-#12 and 5-35/#2-#12 contains 274 values below the RECREL(5-1,5-35) value, 53.

5.8 SUMMARY

The existence of a given relation between a pair of SCs establishes a distinct type of similarity between them. This type of similarity may correlate closely with the one whose existence we can demonstrate with a similarity measure. A general tendency does not guarantee that the correlation exists in every individual case, however. At times the correlation may be far from evident, and in some cases it may fluctuate between strong and weak.

In other words, it seems that different aspects of SC similarity can sometimes exist in unison and sometimes conflict with one another. Due to the dynamic nature of musical materials, structures and processes, no universally applicable hierarchy can exist between these aspects. The validity of one can be neither guaranteed nor denied by another.

■ CHAPTER 6
RECREL AS AN ANALYTICAL TOOL
ASPECTS OF ARNOLD SCHÖNBERG'S OPUS 11, NUMBER 1.

6.1 INTRODUCTION

The purpose of this chapter is to examine the RECREL similarity measure as an analytical tool. The work to be investigated is the first of the Three Piano Pieces, Opus 11, (1909), by Arnold Schönberg.

In section 6.2 we will first discuss a number of aspects relevant to pcset-theoretical analysis of atonal music in general, and the present analysis in particular. Among these are segmentation, the use of a similarity measure in analysis, analytical methods and objectives, identification of the specific SC materials to be examined, degrees of SC similarity which can be considered "analytically interesting," etc.

The analysis itself is carried out in three parts. In each part, we will select a distinct viewpoint on the SC structure of Op. 11, No. 1, and examine how it can be described with the help of RECREL. In the first part, in section 6.3, we study the RECREL characteristics of the SC materials found to be structurally important in two earlier analyses. We also see whether RECREL supports the conclusions of these analyses.

In the second part, in section 6.4, we analyse what we will call *palindromic SC successions*, or palindromes for short. We propose that the music contains a number of successions of SCs designed so that the succession is the same from beginning to end and end to beginning. The lengths of these palindromes vary from rather short to relatively long. The textural appearances of the counterpart segments (having the same SC identity and the same distance from the axis or center point) vary from practically identical to highly different, depending on the case and the positions they have within the palindrome.

In the third part, in section 6.5, we devise a type of SC complex or family, to be called a *RECREL region*, and investigate aspects of the SC structure with the help of such region. In constructing a region, a reference SC, which we will call a *nexus* SC after the terminology in Forte (1973a:101), is selected and a RECREL value representing the highest acceptable degree of dissimilarity determined. To qualify as a member of a given region, a SC has to have a RECREL value with the nexus which is below or at the limit. We believe that the type of harmonic structuring that can be identified and described with the help of the RECREL region concept is an intentional element of the music. There is even a passage containing an interplay between two simultaneous regions.

Besides observing these compositional approaches, we shall also observe combinations of them. A palindromic SC succession may coincide with material described with a RECREL region, for example, resulting in a multi-layered passage which processes several independent elements simultaneously. These observations are in agreement with aspects of harmonic, textural and motivic richness analysts have shown in atonal Schönberg.[1]

6.2 ANALYTICAL PRELIMINARIES

Isaacson summarizes the various aspects relevant to a pcset-theoretical analysis of atonal music. Segmentation is vitally important, since the grouping of pitches in the music is the foundation on which an analysis rests. In spite of many attempts to define a segmentation methodology, however, no general agreement has been found. Even rigorously formal approaches can end up following ad hoc guidelines based on musical intuition. Many aspects have an effect on segmentational decisions, both alone or in combination: rests, instrumentation, phrase markings, articulation, dynamics, attack and release points, texture, register, and rhythm. Yet segmentation based on careful observation of surface features may not be sufficient on its own. Concealed SCs, as well as interconnections between them, can be of great structural significance in a piece of music. When using a similarity measure and identifying potential patterns of SC similarity within a

[1] See, for example, Forte (1972) and Christensen (1987). Even Schönberg's extra-musical activities bear wittness to his tendency to multidimensional thinking: he devised a game of three-dimensional chess.

piece, a number of alternatives present themselves: Are the SC materials of a given piece homogeneous as a whole, or do they contain contrasting elements?; Within a section of music, can a pattern of similarity be described in terms of segment-to-segment similarity between consecutive segments, or in terms of similarity to some nexus SC?; If such a pattern is identified, does it recur elsewhere in the music?; What is the relation between horizontal and vertical dimensions?; etc. (after Isaacson 1992:194-201).

When using a similarity measure in analysis, help in the form of previously accumulated experience is not at hand. A well-established corpus of large-scale analyses using the measures does not exist. As Isaacson points out in (1992:193), when discussing their measures, theorists offer either small-scale analytical examples, or none at all.

Profiling the present analysis of Opus 11, Number 1 was reasonably straightforward, however, for two reasons. Firstly, the basic features suggested themselves already during initial exploration of the music and became the central elements of the analysis, guiding further segmentation. Secondly, part of the task consisted of examining results demonstrated by others. The starting point was that the main conclusions in Wittlich (1974) and Forte (1982), stating that relatively limited groups of SCs constitute the structural basis of the music, are valid.

As the topics to be discussed are very different from each other, it is not meaningful to try to define a distinct set of notions constituting our analytical method. Simply, we identified patterns of harmonic processing - some with the help of RECREL, some without - and became convinced of their importance in the music. Our analytical objectives, then, are to describe these patterns and assess how RECREL can help us in doing this.

The diversity of our topics is particularly obvious with respect to segmentation. In this sense the palindromic SC successions and RECREL regions are each other's exact opposites. In order to be credible, a SC palindrome must be unambiguously segmented, each SC in the succession matching its counterpart at the other side of the center point. A RECREL region, by contrast, manifests its presence in a large number of segments, both obvious and veiled, constituting the nexus SC or classes similar to it. Because of these differences, we will not define any general segmentational guidelines. Whenever necessary, segmentation will be discussed in the context of the current viewpoint.

Opus 11, Number 1, dubbed Schönberg's "first atonal masterwork" by

Forte in (1982), is discussed in a wide range of texts.[2] The present analysis does not attempt to be a comprehensive analysis of the music, not even with respect to pitch organization. We will concentrate on the SC structure, especially on SCs of cardinalities 4, 5, 6 and 7. We believe that SCs of these cardinalities are structurally the most important in this piece.

6.2.1 Determining a "Low" Value

When an analysis is prepared with the help of a similarity measure, and especially with the help of concepts like RECREL regions using user-definable value limits, an important question inevitably emerges: What sort of values are "low" in terms of being analytically significant in a given context? Obviously, a universal value limit below which everything would be "similar" and above "dissimilar" does not exist. Analytical situations, as well as opinions describing those situations, vary strongly. A value limit should be thought of only as an approximate "rule of thumb" -type of concept.

In the analysis below, value limits of the RECREL regions lie between 20 and 25. A limit will not be applied rigorously. Pairs of segments with even considerably higher values will be discussed if their relation is considered important. Two notions led us to identify the value range 20-25 as suitable for our purposes. The first one was the analysis of the RECREL values that belong to pairs of inversionally related and Z-related SCs, pairs of inclusion-related SCs of cardinalities n and n-1, as well as to pairs enjoying other relations often associated with close similarity. The second notion consisted of aural assessments of chord pairs derived from SCs with different RECREL values.

The first notion is obvious. If we have a relation giving rise to associations of similarity, or in some cases even considerations of equivalence, the RECREL values of the pairs enjoying the relation might reflect this special closeness by being "low."[3] We saw in chapter 5 that the values between inversionally related classes vary from 0 to 25. The average is 14. Among Z-related pairs the corresponding figures are 8, 27 and 20, among non-self-complementing 6-pc complement pairs 13, 27 and 22. The average value between a SC of cardinality n and its subset-classes of cardinality n-1 is 27, and so on. On the basis of these figures the

---

[2] For a survey of some of these, see for example Brinkmann (1969:40-57) and Forte (1982:129-30).

[3] As was already noted in chapter 5, however, we cannot draw straight parallels between RECREL similarity and, for example, inversional similarity. The results are only indicative.

lower end of our suggested limit range, 20, seems in fact somewhat restrictive.[4]

During the aural assessments, chord pairs with certain similar characteristics were derived from SC pairs having values from 0 to 100. In our experience there was a correlation between a low RECREL value and chordal similarity. Chords derived from SC pairs with RECREL values approximately up to 20 seemed to cause an impression of similarity, whereas chords from SC pairs with high values did not. This is a purely subjective assessment and is not offered as proof of the validity of the value limit. No controlled listener group tests were conducted to seek confirmation of this observation.[5]

6.3 RECREL AND SC MATERIALS IN TWO EARLIER ANALYSES

Both Wittlich (1974) and Forte (1982) suggest that the harmonic organization of Op. 11, No. 1 is based on a relatively small collection of SCs. In Wittlich (1974:43) the collection is called the Most Prominent Sets, in Forte (1982:132-5) The Harmonic Vocabulary (henceforth just Vocabulary). The two collections differ somewhat, but have also a number of common members. Most Prominent Sets contains 3-, 4- and 6-pc classes, with 5-pc formations being treated as subset-classes of the 6-pc classes. The Vocabulary, for its part, contains SCs of cardinalities from 4 to 8. Each member also has its complement class present in the collection.[6]

According to the two studies, not only are many clearly profiled melodic and chordal formations in the music derived from members of the collections,

---

[4] Frequencies of values up to 20 offer an additional viewpoint. In the value group #4/#4, these values constitute the lowest 2% of values. In the value group #5/#5, the corresponding figure is 4; In #6/#6, also 4; In #7/#7, 10; In #8/#8, 33. Values below a given limit will be more frequent when segments constituting large SCs prevail, an aspect to be taken into consideration when the limit is determined.

[5] To derive chords from some SC X, all permutations were first made from its prime form. The resulting group of ordered pcsets was interpreted as a group of chords so that in each pcset, the order of the pcs from left to right became the order from bottom to top, and the distance between consecutive pitches did not exceed a major seventh. The chords were placed in middle register. Given two groups of chords derived from SCs X and Y, a referential X-chord was determined and a Y-chord selected which best fulfilled the following conditions: (1) its width had to be equal to that of the X-chord, (2) the number of pitches it had in common with the X-chord had to be equal to the largest number of pcs in common between any two member sets of X and Y, (3) the distance between non-common pitches was to be as small as possible. If conditions 1 or 2 could not be met, the X-chord was changed. Transposition was allowed.

[6] The Vocabulary set-classes, each with its complement, are as follows: 6-Z3/6-Z36, 6-Z10/6-Z39, 6-Z13/6-Z42, 6-16, 6-Z19/6-Z44, 6-21, 5-13/7-13, 5-Z17/7-Z17, 5-Z18/7-Z18, 5-21/7-21, 5-Z37/7-Z37, 5-Z38/7-Z38, 4-7/8-7, 4-19/8-19. We will examine the classes under T_n-classification, increasing their number to 44.

but instances of the SCs are constituted also from a wide range of other pitch combinations. Fragments of simultaneous, overlapping or successive principal instances together produce more ambiguous secondary ones, resulting in an extraordinarily skilful fabric of basic materials, a sort of musical cross-word puzzle. Or, in the words used by Forte (1982), a magical kaleidoscope.

Wittlich relates SCs in the Most Prominent Sets by comparing subset-class relations and ICVs. Also verbal assessments are given, for example deeming a SC pair "totally dissimilar." (1974:44-5). Usually the focus is on aspects other than SC similarity, such as pitch relations. Forte discusses SC similarity more often, but does not analyse the Vocabulary with his own similarity relations. It becomes obvious, however, that he sees it as containing a very wide spectrum of harmonic resources. He states, for example, that the composer selected hexad classes of "considerable diversity as well as similarity with respect to interval-class content," and that this aspect has an effect on the sonic richness of the work (1982:135).[7]

Analysis of these SC collections with RECREL strongly supports the notion of harmonic diversity. Let us first make observations at an abstract level, momentarily disregarding the way the classes are actually used in the music. As the collection given by Forte is larger of the two, we will use it as our point of reference. Concerning the relation between the Vocabulary and RECREL, the most important result is this: almost all Vocabulary classes produce very low values with at least a few others. In principle, a succession of Vocabulary class statements could be formulated so that each statement is preceded, paralleled and followed by statements highly similar to it. On the other hand, there are no Vocabulary classes having only low or even moderately low values with the rest of the SCs. The possibility of using contrasting Vocabulary material arrangements is also retained.

The average RECREL value of all pairwise Vocabulary class comparisons is 31. On average, the lowest value obtained when a Vocabulary class is compared with all other Vocabulary classes is 14. The corresponding highest value is 45. The lowest value from an individual comparison is 4, belonging to the pair {5-21A,5-21B}. The highest value, 52, belongs to a few pairs, {4-19A,6-Z42} among them. Interestingly, for some set-classes the *lowest* values are about the same as our RECREL region value limits. The lowest values 6-Z13 and 6-Z42 produce with other Vocabulary classes is 22. 6-16A and 6-16B have almost as high

---

[7] Teitelbaum offers a related observation in (1965:106-8). When certain works of Schönberg and Webern were analysed and compared, the harmonic materials used by the former were found out to be clearly more diverse than those used by the latter.

minimums, 20. 6-21A and 6-21B, in turn, have very few low values. The lowest is RECREL(6-21A,6-21B) = 14. The next lowest value they produce is 25, 6-21A with 5-13B and 7-13B, 6-21B with the inversional classes of the latter two. The rest of the values are 29 or above.

According to Forte, the way the composer uses the Vocabulary in the music does not aim at harmonic unity, rather the opposite, "kaleidoscopic" arrangements bringing out the different dimensions of the SC materials in a multitude of prismatic reflections. He states: "the harmonic structure of the music is constantly in flux, constantly shifting, to reveal new facets of the interlocking of its components." (1982:140). RECREL supports this conclusion as well. The manner with which the Vocabulary classes are usually arranged in the music does not produce a pattern of consistently low RECREL values - or consistently high ones, for that matter. Values between juxtaposed classes routinely vary from low ones (15 or less) to quite high ones (40 or more). A succession of few low values may be found, only to be interrupted by a sudden leap to much higher ones. However, even if harmonic diversity were the rule in arranging the Vocabulary materials, there are still exceptions. In some passages, such as in the closing measures, classes similar to each other are dominant to the extent one must assume a Vocabulary class arrangement intentionally aimed at a uniform harmonic profile.[8]

In the following we shall take an example of both types of Vocabulary class arrangements. The first one describes the norm, a kaleidoscopic SC arrangement, as it appears in the opening music. (Section 6.3.1). The closing measures exemplify the exception, an arrangement with uniform harmonic characteristics. (Section 6.3.2). Many segmentations will be adopted from Forte (1982) and Wittlich (1974), with additional segmentations of our own.[9]

---

[8] The notion of an arrangement like this is based on assessing the Vocabulary profile of a given passage as a whole, with "rule of majority" as the main criterion. It is not required that each and every Vocabulary class instance must be of uniform characteristrics since it is evident, for example, that many similar instances can have one dissimilar in their midst, or that classes of both small and large cardinalities will be present, together producing high RECREL values.

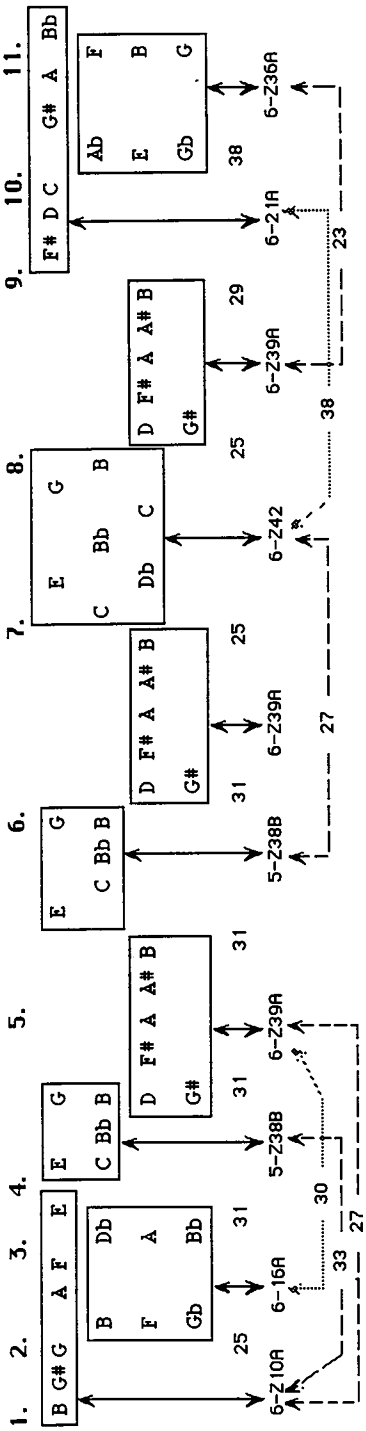
[9] The examples will use diagrammatic notation similar to the one in Forte (1982). Some examples are placed sideways. Numbers between the arrows pointing to the SC names are RECREL values. We will mainly concentrate on RECREL values between adjacent, overlapping and simultaneous segments, but some additional RECREL values may be given below the SC names. These belong to pairs of segments having one or more other segments between them.

6.3.1 The Opening: Measures 1-11

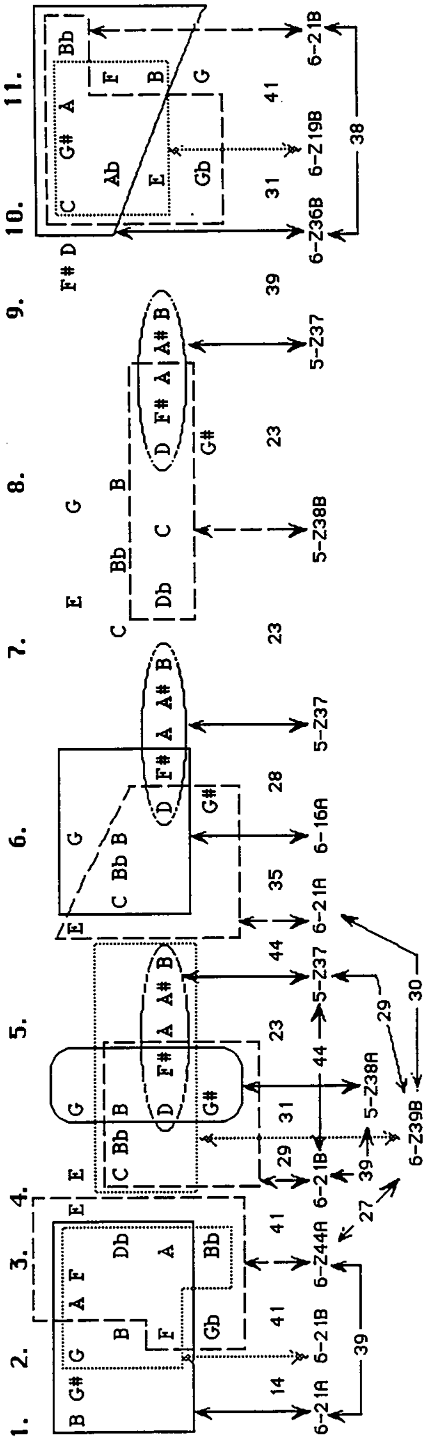
The opening music, mm. 1-11, can be seen in Ex. 6.1. The segmentation is shown in two separate examples, 6.1.a and 6.1.b. The former contains principal segments, the latter more ambiguous ones. Part of this passage, from m. 7 onwards, will be discussed also in connection with a palindromic SC succession. (Section 6.4.4).

EXAMPLE 6.1: Mm. 1-11.

The average RECREL value in Ex. 6.1.a is approximately 30. Most values are between 25 and 31. Here, as well as in Ex. 6.1.b, the highest values belong to SC pairs containing either 6-21A or 6-21B. These inversionally related classes were found above to be close only to each other and rather distant from most of the other Vocabulary classes. This is evident in Ex. 6.1.b. The first value, 14, belongs to the pair {6-21A,6-21B} itself, the following two values, both 41, to the pair {6-21B,6-Z44A}. The average value is approximately 32.



EXAMPLE 6.1.a: Vocabulary class instances in mm. 1-11. Principal segments.

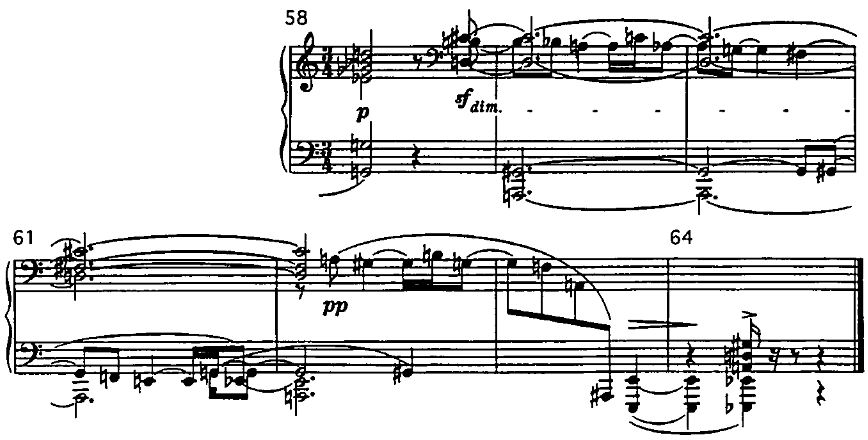


EXAMPLE 6.1.b: Vocabulary class instances in mm. 1-11. Secondary segments.

6.3.2 The Conclusion: End of m. 58 - m. 64

According to Forte (1982:167), multiple forms of SC 6-Z3 pervade the Coda. In our analysis the two inversionally related Vocabulary classes 6-Z3A and 6-Z3B are the components producing the dominant harmonic element of the passage, 7-1. For every member set in 6-Z3A there is a member set in 6-Z3B so that the union of the two constitutes a pcset belonging to this inversionally symmetric "chromatic" class.[10] We believe these "union instances," covering most of the conclusion in many different shapes, are intentional. A single instance could be dismissed as a random formation, but here they occur in a concentration too dense to be mere coincidence.[11]

EXAMPLE 6.2: Mm. 58-64.



Example 6.2.a gives seven 7-1 instances, located in six measures. They consist of linear formations, sustained chords combined with simultaneous pitches in the linear formations, and consecutive attack points. The instance belonging to the

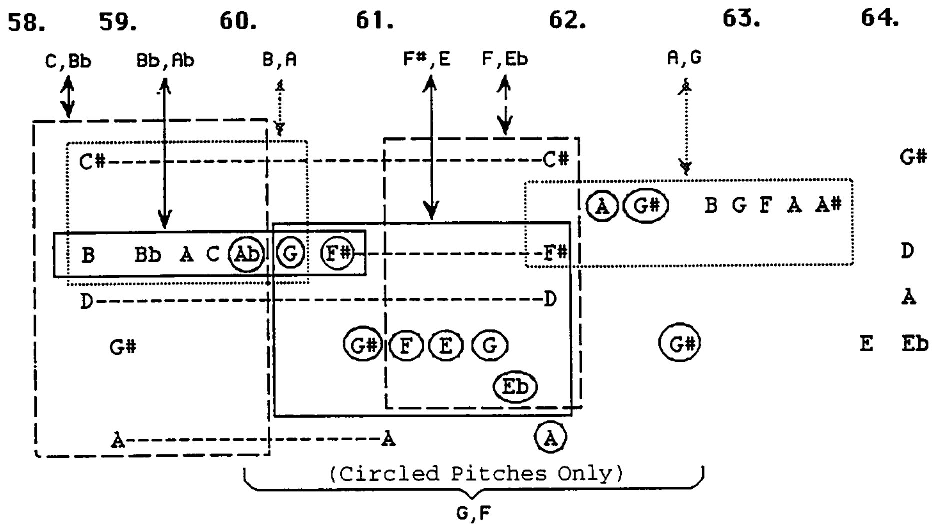
[10] 7-1 is not a Vocabulary member itself. As 6-Z3A and 6-Z3B are used as the building blocks of the dominant SC, the conclusion is not the purest possible example of a Vocabulary class arrangement, of course. The passage was chosen because it fulfils its purpose well, illustrating an arrangement of uniform harmonic characteristics.

[11] The juxtaposition of inversionally related classes is an aspect occurring throughout the movement. Compare, for example, segments within and between the examples 6.1.a and 6.1.b. Overlapping segments constitute inversionally related pairs 6-Z39A/6-Z39B, 5-Z38A/5-Z38B, 6-21A/6-21B and 6-Z36A/6-Z36B. Example 20b in Forte (1982:166) gives additional examples from the passage under discussion. The pairs are 6-Z10A/6-Z10B and 6-21A/6-21B. The former constitutes 7-Z37, itself a Vocabulary member, the latter 7-8.

last category, marked with circled pitches, covers half of the passage. Each instance is associated with a pair of pcs, containing the elements to be omitted in order to obtain the subset-classes constituting 6-Z3A and 6-Z3B, respectively. The RECREL value between 6-Z3A and 6-Z3B is 14, between 7-1 and 6-Z3A (and 6-Z3B) 17.

The last chord Eb-A-D-G# constitutes an instance of SC 4-9, with the preceding E an instance of 5-7A. These classes differ completely from their harmonic surroundings. RECREL(4-9,7-1) = 60, RECREL(5-7A,7-1) = 48. For a discussion of this striking contrast, see Forte (1982:167), Wittlich (1974:52) and Perle (1981:12).

EXAMPLE 6.2.a: End of m. 58 - m. 64. Instances of SC 7-1. Pairs of pcs above arrows indicate which elements must be excluded in order to get instances of 6-Z3A and 6-Z3B, respectively



6.4 PALINDROMIC SC SUCCESSIONS

Palindromic, or more generally, symmetric arrangements of musical objects (pitch and chord successions, textures, even entire works) are notions relevant to many 20th century works and theoretical approaches. The status of such an arrangement may be anything from an all-important structural principle to a passing detail.[12]

12 See, for example, Neumeyer (1986:228-38) and Isaacson (1992:211), discussing Hindemith's *Ludus Tonalis* and Webern's Op. 5, No.2, respectively. Cherlin (1991) gives an interesting analysis on palindromes in general and interval palindromes in Schönberg's *Moses und Aron* in particular, providing also a list of sources discussing "inversional balances" in Schönberg's music (Ibid., 69-70, *n* 3).

The existence of palindromic arrangements in Op. 11, No. 1, then, is not surprising as such. There is an element of special interest in them, however, i.e., the objects being arranged: formations whose pitch contents, registral positions, textural profiles, etc., may vary, but whose SC identities show a pairwise correspondence around a center of symmetry.[13] An argument supporting the intentionality of the palindromic SC successions is obvious: it is highly unlikely that several such arrangements would be formed accidentally. We believe they indicate how naturally the composer was thinking in terms of the abstract musical entities we today call set-classes.

It seems reasonable to assume that the SC palindromes originate in more obvious palindromic arrangements, with evident textural, pitch-to-pitch or intervallic correspondence. Lessened resemblance between counterpart classes was then perhaps allowed, as long as their SC identities were maintained.[14]

The number of palindromic SC successions to be examined below is four. They are not in the order they appear in the music. The most obvious cases will be discussed first.

6.4.1 Measures 34-38

The music is shown in Ex. 6.8 in section 6.5.2.2 below. The most obvious element in the passage is the succession of parallel thirds, forming the upper strand of the right-hand part. The first six thirds constitute a modified restatement of the opening theme.

Ex. 6.3 gives the palindromic layer of the music. The succession of thirds is divided here into four segments, of which the first, third and fourth constitute instances of 5-21A and the second an instance of its inversionally related class,

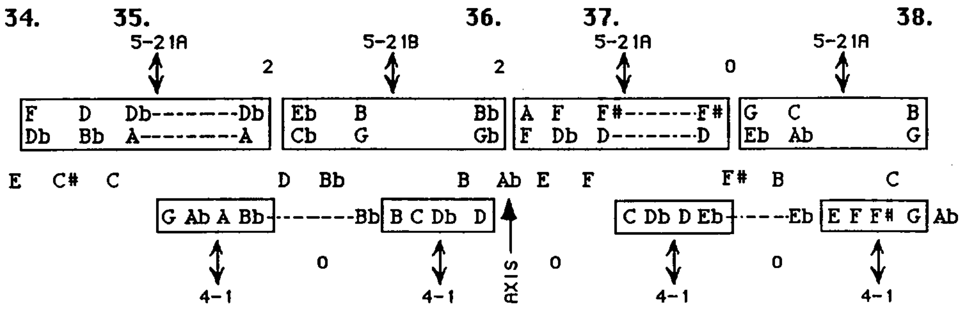
---

[13] Palindromic arrangements are not the sole property of the SC structure, of course. An example belonging to another domain is the textural palindrome in the opening measures. (Examples 6.1 and 6.1.a) The right-hand 6-Z10A statement in mm. 1-3 has a counterpart in the melodic 6-21A statement in mm. 9-11; The harmonic 6-16A statement, laid out in two trichords in mm. 2-3, corresponds to the related 6-Z36A combination in mm. 10-11, etc. See Wittlich (1974:45), Perle (1977:162-3) and Forte (1982:142). Note also the observation in Wittlich (1974:52) on the symmetric arrangement of certain instances of SCs in Most Prominent Sets. The transpositional levels of counterpart segments correspond, suggesting another aspect of palindromic or symmetric processing.

[14] Schönberg's own comments on different aspects of mirror arrangements in music tell of varying levels of appreciation, and do not suggest what his attitude or approach might have been towards the idea of palindromic SC successions. In (1975:68-9) he scorns his mirror-canon exercises as "not music but only gymnastics." In (Ibid., 220-3), by contrast, he discusses the use of "mirror forms" of a motif as means of creating thematic cohesion, stating that it does not matter whether or not the forms are adopted intentionally. They can also be "a subconsciously received gift from the Supreme Commander."

5-21B.[15] The regularity of the segments is exceptional. They do not overlap, they do not exclude any thirds in the succession, and together they cover exactly the whole passage. Each 5-21 instance is paired with an instance of SC 4-1 in the left-hand part. 4-1 is not a Vocabulary member but belongs to the Most Prominent Sets in Wittlich (1974).

EXAMPLE 6.3: Mm. 34-38. A palindromic SC succession.



The pair {5-21B,5-21A} immediately around the axis is one of just two cases where palindromic counterparts are not instances of the same class, but of inversionally related classes. The other case consists of classes 4-19A/4-19B in the mm. 7-24 palindrome. (Section 6.4.4). In all other cases the counterpart classes represent the same transpositional SC. Interestingly, the pairs in both cases have exceptionally low RECREL values, as if to suggest that substituting one with the other is a minimal deviation from the usual regularity. RECREL(5-21A,5-21B) = 2; RECREL(4-19A,4-19B) = 8. The right-hand and left-hand strands are consistent internally but highly different from one another. RECREL(4-1,5-21A) = RECREL(4-1,5-21B) = 68.

Mm. 34-38 will be analysed also in terms of a RECREL region in section 6.5.2.2. For analysis of the Vocabulary materials, see Forte (1982:158-60).

6.4.2 Measures 55-58

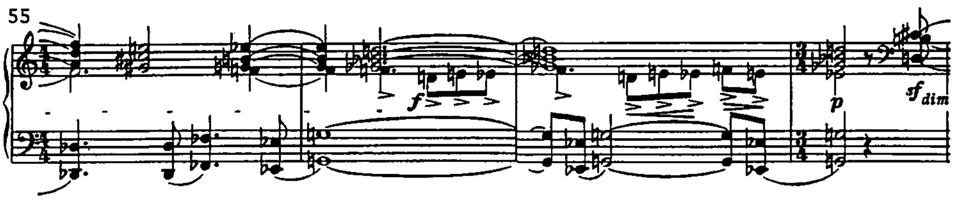
The music and segmentation are given in examples 6.4 and 6.4.a, respectively. The palindrome consists of 11 classes, four of which also contain subset-classes.

---

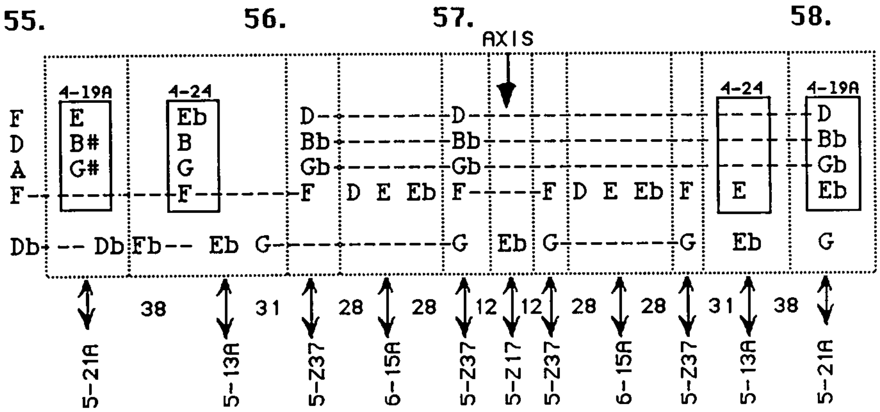
[15] The multiple instances of 5-21 are mentioned also in Forte (1982:159-60).

The subset-classes constitute either 4-19A or 4-24.[16] The axis class is the instance of 5-Z17 in m. 57.

EXAMPLE 6.4: Mm. 55-58.



EXAMPLE 6.4.a: Mm. 55-58. A palindromic SC succession.



The counterpart classes are texturally highly similar in the center area, less so in the outer areas. Low RECREL values between consecutive segments are to be found, but only among clearly higher ones. With the exception of SC 6-15A and 4-24 all classes are Vocabulary members.

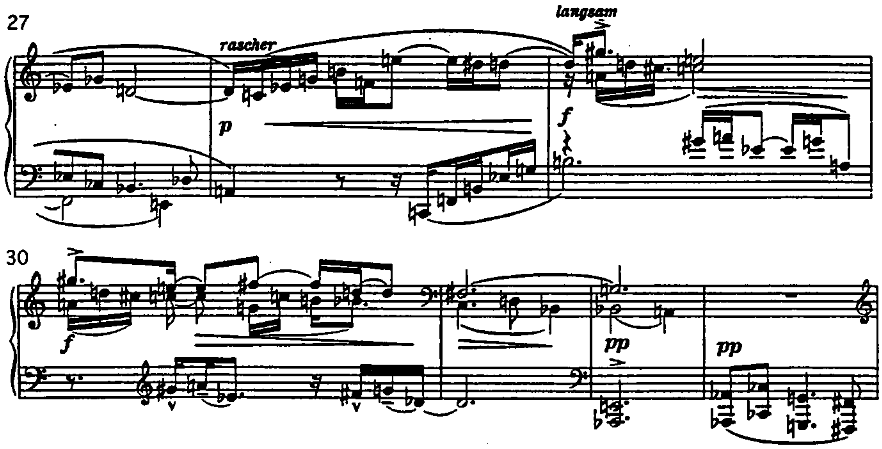
6.4.3 End of m. 27 - m. 33.

The number of classes in this palindromic SC succession is eleven, several small formations being embedded in larger ones. Notably none of the classes are Vocabulary members. (Ex. 6.5.a). SCs in the axis area again bear clear textural re-

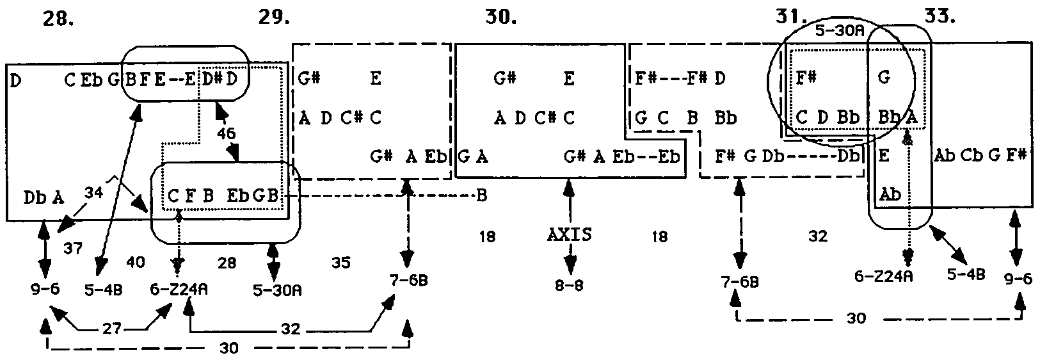
16 The pitch identity of the middle element in the second right-hand chord in m. 55 is in dispute. Perle (1981:15 *n* 4) suggests B#, Forte (1982:165 *n* 18) C#. We adopt the former alternative.

semblances to one another. The axis class 8-8 and the two instances of 7-6B around it (mm. 29-30) are almost identical. The outer areas, in contrast, are completely different.

EXAMPLE 6.5: Mm. 27-33.



EXAMPLE 6.5.a: End of m. 27 - m. 33. A palindromic SC succession.



According to the Overview of Form in Forte (1982:131), the succession covers two distinct passages, "Canonic episode related to 12-16" and "Final statement of a." Moreover, the three pcs in the first 9-6 instance, D-Db-A, form a link with the previous passage. Links between successive passages are present also elsewhere in the music. See, for example, the melodic line constituting an instance of 6-21A in mm. 16-19 of Ex. 6.6.a. (A horizontal segment drawn with a dashed line).

Two low RECREL values are to be seen, between the axis class 8-8 and the

7-6B classes around it. RECREL(8-8,7-6B) = 18. The rest of the values are relatively high.

6.4.4 Measures 7-24

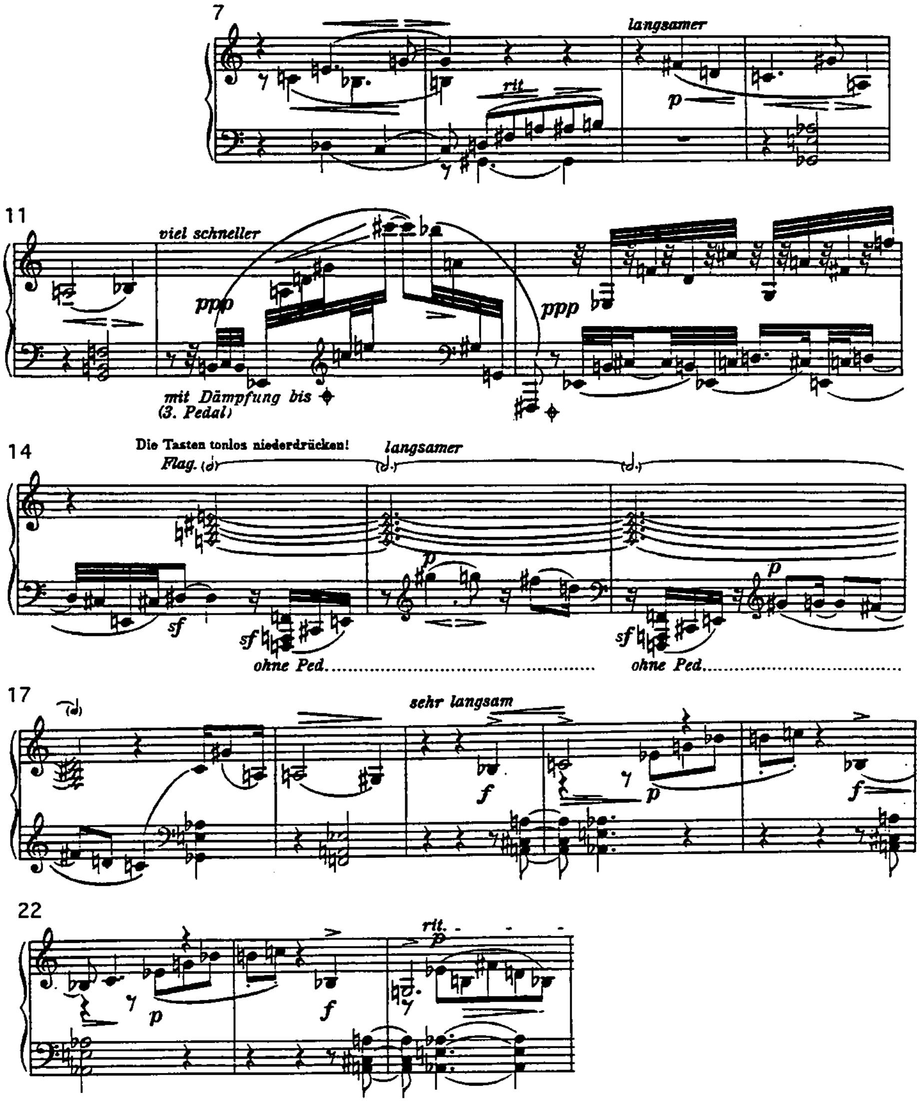
The longest of the palindromic SC successions covers 18 measures, mm. 7-24. The number of classes is 24. (Ex. 6.6.a). The axis is in m. 13, between the two pairs of instances of 4-2A. A number of the SCs, notably the ones in the outer areas, belong to the Vocabulary (6-Z42, 5-Z38A, 6-Z39A, 5-Z37, 6-21A, 4-19A and 4-19B). Most classes in the center area, by contrast, do not (3-6, 3-8A, 8-4B, 5-1 and 4-2A). 3-8A and 4-2A, however, belong to the Most Prominent Sets of Wittlich (1974), as does the complement of 8-4B, 4-4A.

There are two small irregularities in the palindrome. The first one is the order of the two pairs of trichords, constituting 3-6 and 3-8A. (Mm. 10-11 and 17-18). In order for the symmetry to be accurate, the order of one or other of the pairs should be reversed. The other irregularity, between the counterpart classes 4-19A and 4-19B in m. 12 and m.14, respectively, was already being mentioned in section 6.4.1 above.

The palindrome is significantly independent of the formal structure of the movement. In terms of the Overview of Form in Forte (1982:131), it begins in the middle of the b passage in the first A in the Exposition, spans over a' in mm. 9-11, the "new material episode" in mm. 12-14, transition in mm. 15-16, return of a' in mm. 17-18 and return of b in 19-24. The degree of textural similarity between counterpart classes fluctuates strongly. The first and last classes are texturally remote, whereas the two instances of 5-Z37, D-F#-A-A#-B in m. 8 and Eb-G-Bb-B-C in mm. 20-21, are identical, only a minor ninth apart. The latter segment, however, is only a part of a long 5-Z37 statement spanning from m. 19 to m. 24. The corresponding 6-21A statements (mm. 9-11 and 16-19) have highly similar contours and pitch contents, and the right-hand statements of 4-2A, around the axis in m. 13, are identical, save for their transpositional levels.

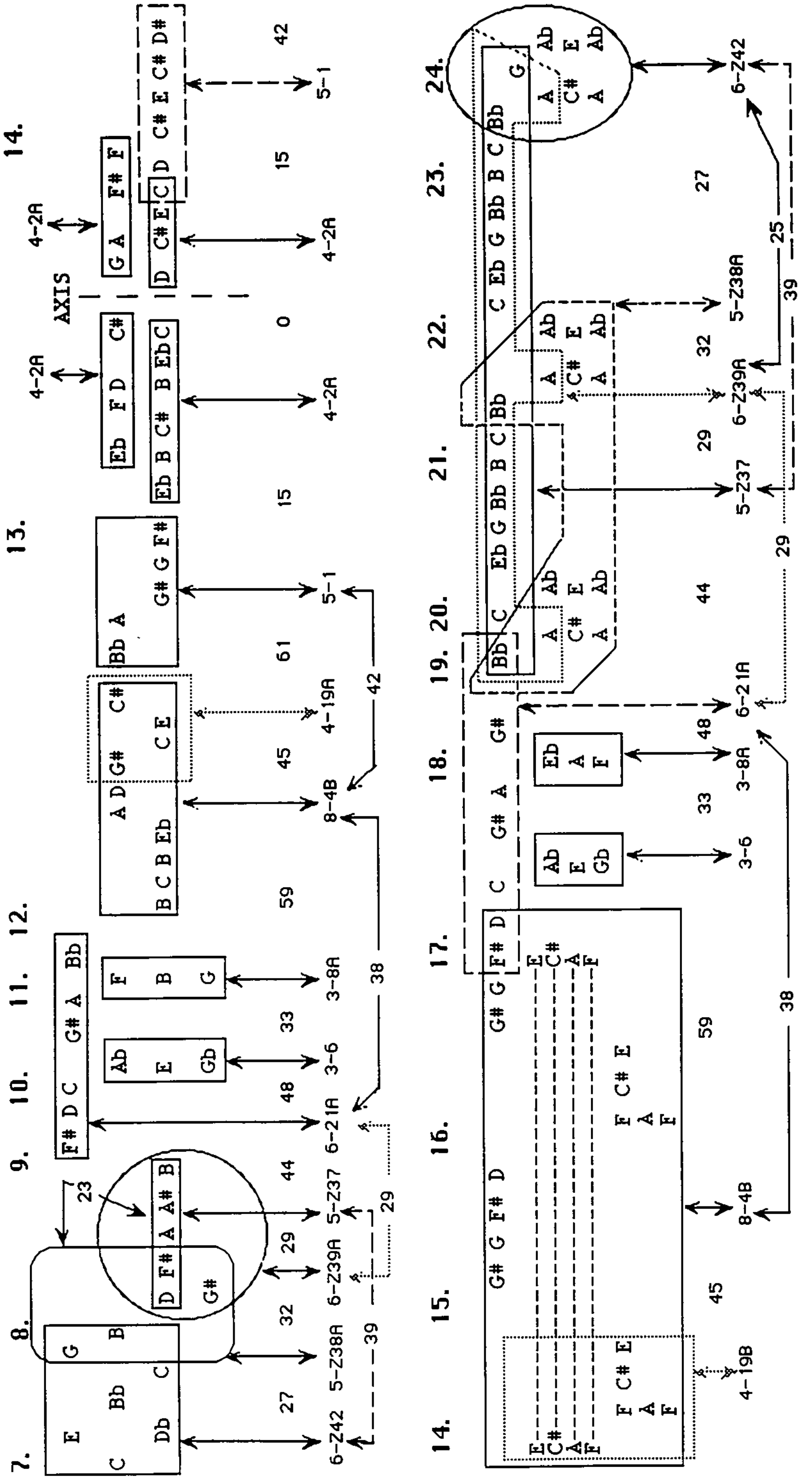
The most intriguing textural differences are to be found between the largest counterpart classes, the two instances of 8-4B. The first one constitutes the arpeggiated pattern in m. 12, (from B to C#), the latter one the sustained F-A-C#-E chord (4-19B) with the twice repeated succession G#-G-F#-D. The length of the first 8-4B instance is less than three beats, that of the second three measures.

EXAMPLE 6.6: Mm. 7-24.



The RECREL values vary considerably. The lowest ones are again found in the axis area, between segments having textural similarities as well. The closing measures of the palindrome, mm. 19-24, contribute to a passage which will be analysed in terms of a RECREL region. See section 6.5.2.1. For details on the Vocabulary and the Most Prominent Sets materials, see Forte (1982:139-51), and Wittlich (1974:45-8).

EXAMPLE 6.6.a: Mm. 7-24. A palindromic SC succession.



6.4.4.1 Some Resemblances between mm. 14-17 and mm. 55-58

It was noted above that SC palindromes are not the only aspects of symmetry in the movement, as textural, transpositional etc., considerations arise as well. Resemblances between mm. 14-17 and mm. 55-58 provide an additional viewpoint. See examples 6.4 and 6.6.

The chord F-A-C#-E in m. 14 belongs to 4-19B. With the pitch D# immediately below it constitutes SC 5-13B, with G# (m. 15) 5-21B, with G 5-26B, with F# 5-Z37 and with D 5-Z17. In mm. 56-57, in turn, the sustained chord G-Gb-Bb-D belongs to SC 4-19A. With the pitch F (m. 56, second beat) it constitutes 5-Z37, with E 5-26A and with Eb 5-21A. If we add to these the instance of 5-13A in the end of m. 55 and the axis class 5-Z17, we see that mm. 14-17 and mm. 55-57 consist of two collections of SCs such that each class in one has its inversional counterpart in the other. (The inversionally symmetric classes being their own counterparts, of course).[17]

Also, the two SC collections are placed in textural settings that are themselves "inversions" of each other: in mm. 55-58 the sustained chord is above the melodic movement, in mm. 14-17 it is below. (In the latter case the chord shares its pc contents with the low left-hand pattern). It is especially interesting that the inversional counterparts in the two collections are not mirror-inverted *chords*, but inversional with respect to *class* identity.

6.4.5 Conclusions: RECREL and the Palindromic SC Successions

On the basis of the examples above, it seems that SC similarity is not a governing principle in the palindromic SC successions. RECREL values of successive classes fluctuate strongly from low to relatively high. The areas around the axes, however, seem to be an exception. Low values prevail, and the segments are often considerably similar also texturally. The consistency with which textural similarities occur in the center areas suggests that the palindromic successions were not meant to be completely concealed.

While one palindrome may exactly cover a distinct passage, another may

[17] In the original segmentation of the mm. 55-58 palindrome, given in Ex. 6.4.a, some of the 4- and 5-pc segments were grouped together to form the two counterpart segments constituting 6-15A. Also this hexad class has its inversional counterpart in the first palindrome: the F-A-C#-E chord in m. 15, taken with the pitches G# and G above it, constitute 6-15B.

extend over several passages. In the latter case, the beginning and end of the palindrome may or may not coincide with changes in texture.

6.5 RECREL REGIONS

The fact that RECREL region membership is limited with an artificially determined value limit makes the concept more ambiguous than the SC families discussed in Isaacson (1992:157-91). The RECREL regions are seen here as purely pragmatic means to filter the examined SC materials. The disqualification of a certain SC from a region does not necessarily imply that it is analytically uninteresting from the point of view of the region.

The RECREL region concept emerged during computerized segmentation of the pitch material. (Details in section 6.5.1 below). In some passages large numbers of consecutive, overlapping and/or simultaneous segments constituted SCs which have low RECREL values with each other. Some of the segments were texturally obvious entities (chords, phrases, etc.), others just secondary "clouds" of pitches residing near each other. An important notion was also that the segments covered the passages completely. After further analysis it became obvious that the best way to describe this type of cohesion is with a reference harmony, a nexus SC. It also soon appeared that the type of harmonic processing to be described with the help of the RECREL regions was an independent layer in the music. Instances of the Vocabulary classes were present in the music, but arranged for example so that some of their pitches contribute to one RECREL region segment, the rest to another. A Vocabulary class can participate in a RECREL region, of course, but it can also be very distant from the nexus.

Generally, the notion of a complex texture whose different strands combine into different types of harmonic entities has been identified as an element of central importance in atonal Schönberg.[18] The simultaneous Vocabulary, RECREL region and/or SC palindrome layers exemplify this type of thinking in Op. 11, No.1.

18 See Forte (1972). The reader is directed to the especially illuminative example 11 (Ibid., 57). It shows four simultaneous "chordal streams" embedded in a single texture, resulting from different segmentations and producing repeated instances of four SCs. The work under discussion is the fourth movement of the Five Orchestral Pieces, Op. 16.

6.5.1 The Automated Segmentations: Nexus Selection and RECREL Region Validity Conditions

Once initial computer searches had suggested that some passages contain harmonic processing suitable for RECREL region analysis, the entire movement was searched. Only some passages were eventually found to contain such processing.

The data given to the computer consisted of pitches only. Right-hand and left-hand parts were examined both separately and together. Simultaneities were interpreted as successions so that they could be imbricated. Phrasing, rhythm, dynamics, etc., were ignored.[19] To launch a search, a pitch succession representing a passage was segmented into overlapping segments of several different lengths. Typically, the shortest segments were of length 3, the longest of length 16.[20] Even a short passage produced hundreds of segments, SC identities of which were determined. After this, every SC between cardinalities 4 and 8 was set in its turn to be a "nexus candidate." For each candidate, all RECREL values with the segments were found. Only segments whose values with the candidate were under the limit (typically between 20 and 25) were retained. Statistics were compiled for every candidate, by analysing numbers of accepted segments, average RECREL values and segment distribution within the passage.[21]

Passages producing unsatisfactory results were rejected from further testing. Their best nexus candidates could perhaps produce low values, but with segments concentrated in only a part of the passage. The whole material was not covered consistently enough. In accepted passages, however, low-valued segments belonging to a given candidate or a few candidates could cover the whole passage well. Further testing was then done manually, in order to see whether segments outside the scope of the automatic imbrication would corroborate the preliminary results.[22] Determining the final nexus SC was difficult in a few

---

[19] It should be evident without saying that such searches were thought of as producing very crude analytical raw material only.

[20] When segmenting the pitch succession into 4-element units, for example, the first segment contained pitches 1-4, the second one pitches 2-5, the third one pitches 3-6, etc.

[21] Whether or not the nexus itself should be prominently present in a passage analysed in terms of a RECREL region is an interesting question. A passage without *any* instances of the nexus and still conveying a high degree of RECREL region cohesion is not impossible to envisage. This would suggest a sort of harmonic "center of gravity" outside the SC materials of the passage itself. Analogous cases with this kind of "ghost material" feature would perhaps be variations without a theme, examples of which exist.

[22] An interesting observation was that the nexus candidates producing best results had often low RECREL values with each other. They also produced low values with many same segments. Thus,

cases, due to equally strong candidates. A 6-pc candidate, for example, could have low values with many evenly distributed 5-pc and 6-pc segments, whereas a 7-pc candidate could produce equally desirable results with emphasis on 7-pc segments.

The automated segmentations were used only during the initial stages of the analysis, and their main purpose was to identify the passages to be exluded from the RECREL region analysis. Part of these segmentations contribute also to the final analysis, but are complemented by segments identified by other means. The analytical conclusions presented below are not dependent upon the mechanical analysis.

6.5.2 RECREL Region Passages In Opus 11, Number 1

As mentioned above, the type of harmonic processing which we wish to illustrate with the help of the RECREL region concept is not present throughout the music. The passages where it appears are for the most part associated with the Development section.

The number of passages to be analysed in terms of RECREL regions is three. The first one consists of mm. 19 - beginning of 28, its nexus SC being 6-20 {0,1,4,5,8,9}. (Section 6.5.2.1). The second passage, to be examined in section 6.5.2.2, covers measures 34 - beginning of 38. Again, the nexus SC is 6-20. The third and most complex passage begins immediately after the second one, covering the measures 38-48. It will be analysed in terms of two simultaneous RECREL regions, one of horizontal and the other of vertical orientation. The nexi are 6-20 and 8-28, respectively, the prime form of the latter being {0,1,3,4,6,7,9,10}. (Section 6.5.2.3). Together the second and third passages cover most of the Development section. See Overview of Form in Forte (1982:131).

In the RECREL region examples, each segment is associated with a SC name followed by a number. The latter is the RECREL value the segment has with the nexus SC. In the text below, a number in parentheses after the SC name serves the same purpose.

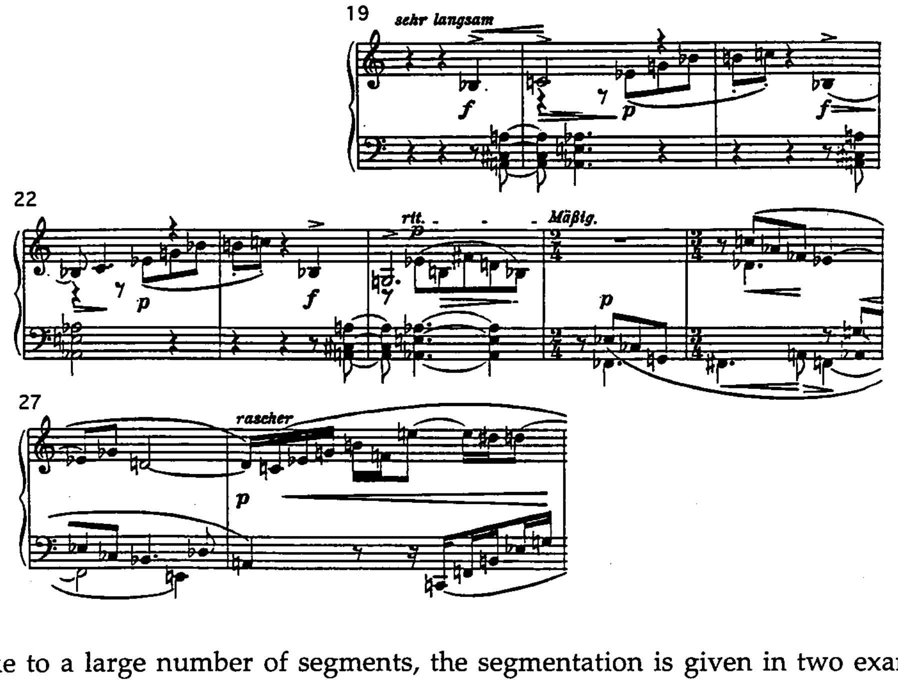
---

the harmonic character of the passage could be identified from different but closely related angles.

6.5.2.1 M. 19 - Beginning of m. 28: Nexus SC 6-20

We noted in section 6.5.1 that, in some passages, more than one SC seemed suitable to serve as the nexus. The present passage is one of them. Both 6-20 and 7-21A {0,1,2,4,5,8,9} suggested themselves strongly during the nexus search. For the sake of brevity we will examine the hexad class nexus only. This inversionally and transpositionally symmetric SC is not a Vocabulary class itself, but has low RECREL values with a number of them.[23] It is the nexus, or one of two nexi, in every passage to be examined in terms of a RECREL region.[24]

EXAMPLE 6.7: Mm. 19-28.



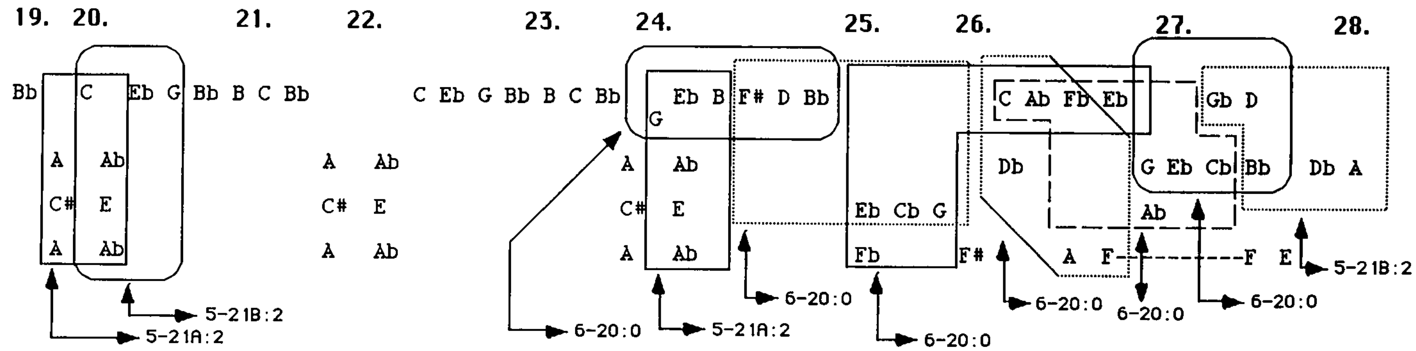
Due to a large number of segments, the segmentation is given in two examples, 6.7.a and 6.7.b. The former contains segments with the lowest values, the latter those with higher ones.

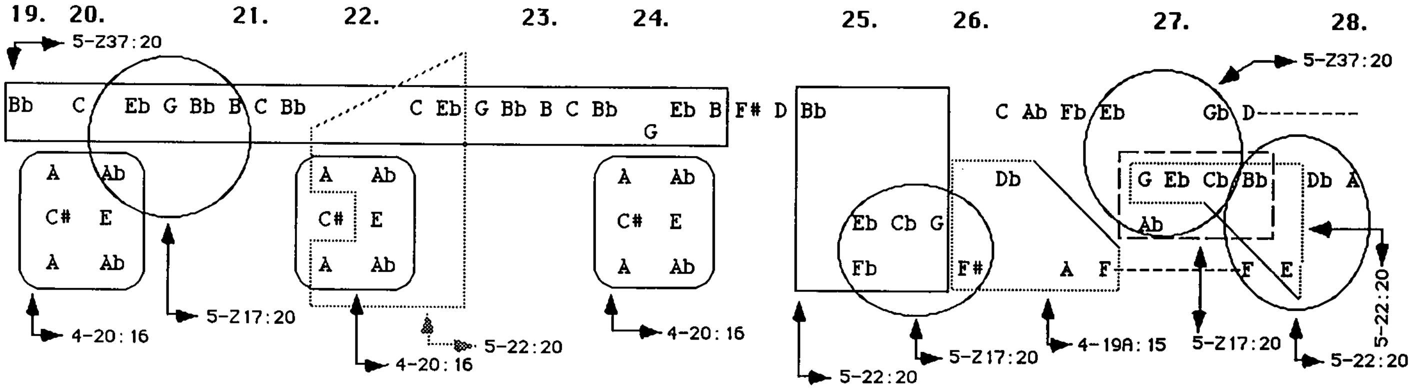
---

[23] The lowest such value, RECREL(6-20,5-21A) = RECREL(6-20,5-21B) = 2, is after the handful of zero values the lowest value in the entire set of RECREL values.

[24] Results shown elsewhere suggest that it might serve well as a nexus also in other atonal works by Schönberg. See examples 6-8 in Forte (1972:52-3). The excerpts are from Op. 20, Op. 19, No. 2, and Op. 22. Given the value limit 20, all 4- and 5-pc SCs in the examples would be members of the RECREL region 6-20.

EXAMPLE 6.7.a: M. 19 - beginning of m. 28. Instances of SCs in RECREL region 6-20. Value limit 20. Segments with lowest values.





EXAMPLE 6.7.b: M. 19 - beginning of m. 28. Instances of SCs in RECREL region 6-20. Value limit 20. Additional segments.

The value limit is twenty. The beginning of the passage has a long melodic line in the right-hand part, the SC being 5-Z37 (20). (Ex. 6.7.b, mm. 19-24). It is followed by the first instance of the nexus itself. (Ex. 6.7.a, m. 24). The left-hand pair of three-note chords, stated three times within mm. 19-24, constitute together 4-20 (16). (Ex. 6.7.b). Different combinations of the chords and the melodic line produce instances of 5-21A (2), 5-21B (2), 5-22 (20) and 5-Z17 (20).

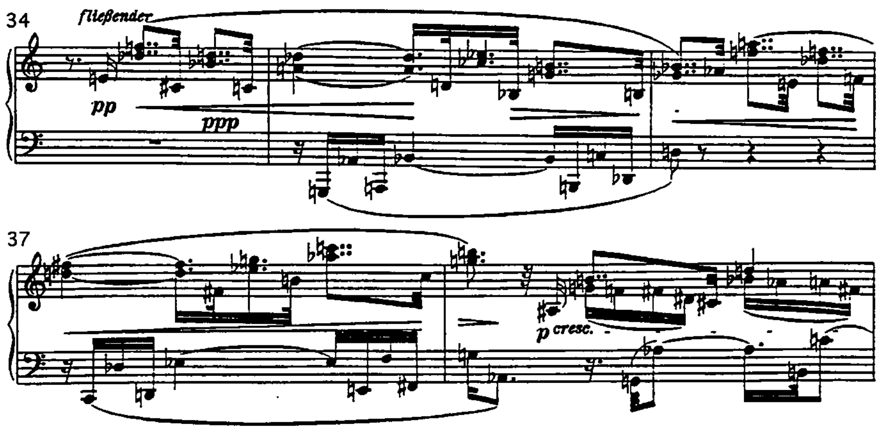
Mm. 24-28 contain six instances of 6-20. They typically overlap so that each of the many melodic "cells" which constitute the augmented triad class 3-12 participates in at least two 6-20 instances. (Ex. 6.7.a). Among the other classes in mm. 24-28 are 5-22 (20), 5-Z17 (20), 4-19A (15) and 5-Z37 (20). (Ex. 6.7.b). The Vocabulary has a strong representation, as 6-20, 4-20 and 5-22 are the only classes not belonging to it.

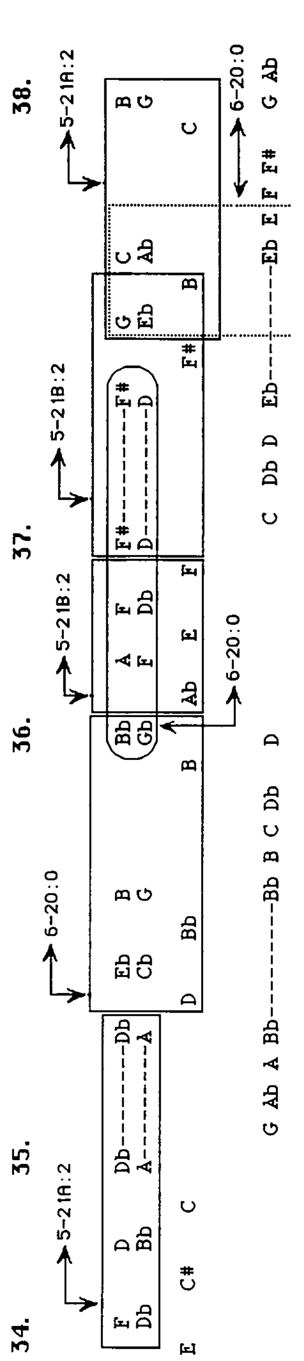
A part of the passage participates also in a palindromic SC succession. See section 6.4.4. For a description of the Vocabulary materials, see Forte (1982:150-4).

6.5.2.2 Measures 34-38: Nexus SC 6-20

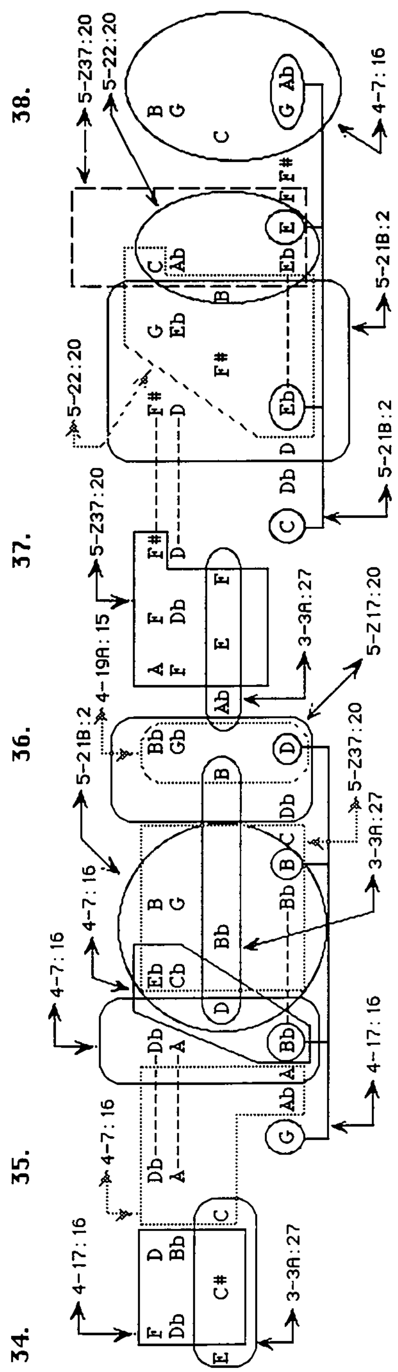
The segmentation is again given in two examples, 6.8.a and 6.8.b. The former contains segments with the lowest values, the latter those with values nearer to the limit, 20. With the exception of the three pitches at the beginning of the passage, the segmentation in Ex. 6.8.a covers the whole right-hand part. One segment also contains two left-hand pitches. The segments constitute SCs 6-20 (0), 5-21A (2) and 5-21B (2).

EXAMPLE 6.8: Mm. 34-38.





EXAMPLE 6.8.a: M. 34 - beginning of m. 38. Instances of SCs in the RECREL region 6-20. Value limit 20. Segments with lowest values.



EXAMPLE 6.8.b: M. 34 - beginning of m. 38. Instances of SCs in the RECREL region 6-20. Value limit 20. Additional segments.

The left-hand part, with its chromatic patterns, seems at first to be in stark contrast with any other dimension in the material. However, it turns out to be tied to the RECREL region in a number of ways. (Ex. 6.8.b). Even if the left-hand chromatic patterns are investigated separately, a textural factor suggests a connection between them and the nexus. When the lowest and highest pitches in each four-note pattern (as well as the very last Ab, in the beginning of m.38) are analysed separately - not a very drastic step considering the emphasis each outer pitch gets in the patterns - they constitute instances of 4-17 (16) in mm. 35-36, and 5-21B (2) in mm. 37-38.

Instances of 3-3A (27) can be seen among the many 4-pc and 5-pc classes. The value is above the limit 20, but was accepted as it is low with respect to the cardinality difference. In the value group #3/#6, 27 belongs to the lowest 1% of values.

The palindromic layer of this passage was examined in section 6.4.1. For a Vocabulary material analysis, see Forte (1982:158-60).

6.5.2.3 Measures 38-48: Simultaneous RECREL Regions 6-20 and 8-28

The question of vertical and horizontal dimensions containing different harmonic materials has been discussed in various sources, both at a general level and specifically with Schönberg in mind. Teitelbaum (1965:112), for example, reports clear differences in horizontal and vertical SC materials in Schönberg's Opus 19. Isaacson (1992:198-9), in turn, refers to this aspect repeatedly when discussing notions that are of relevance to an analysis using a similarity measure. Forte's emphasis is not as much on the differences between the horizontal and vertical materials, as it is on the strictness with which the materials are structured in atonal Schönberg to form the "basic matrix of the music." Each newly composed configuration may affect more than one dimension, creating new SC or completing SCs already partially formed. The result is a strictly controlled, albeit often concealed mixture of multiple dimensions, in contrast to which the surface of the music gives the appearance of utmost flexibility and freedom (Forte 1972:62-3).

Observations like these are consistent with our own regarding the passage under discussion. We believe that the music in mm. 38-48 realizes a harmonic "basic matrix" with three separate strands. At the surface level is the Vocabulary strand, its central elements being the six right-hand patterns constituting 6-16B, in m. 38 and again in mm. 42-43.[25] The RECREL region 6-20 strand, in turn, seems to have the function of governing the transpositional relations between the surface materials. Counterpart pitches in the recurring 6-16B patterns, for example, combine into instances of region 6-20 classes in many different ways.[26] The 8-28 region strand, finally, establishes itself in more "local" segments, consisting of all successive or simultaneous pitches within the assumed segment boundaries. Hence the characterization "vertically oriented" region. The 8-28 region segments are often texturally ambiguous, combining parts of consecutive surface elements, etc.

As analysed by RECREL, the two nexi are distant from each other. RECREL(6-20,8-28) = 48. It was already seen that 6-20 has many low values with the Vocabulary classes. The situation with 8-28 is different. The lowest value it has with a Vocabulary class is 15, with 6-Z13. This hexad class is also its only subset-class among the Vocabulary classes. With 6-Z42, the complement of 6-Z13, 8-28 has the value 25. All other values it has with the Vocabulary classes are between 30 and 51.

Our analysis of mm. 38-48 is in three parts. First we examine the 6-20 region in the measures where it clearly coincides with the 8-28 region. Then we analyse the 8-28 region from the same viewpoint. Finally we investigate excerpts where the functions of the two regions fluctuate. A region may be momentarily non-existent, for example, the remaining one adopting its function during the absence.

6.5.2.3.1 Nexus 6-20

The music to be examined consists of two excerpts, mm. 38-40 and 42-44. (Examples 6.9 and 6.10). Our starting point is the instances of the augmented triad class, 3-12, embedded in the music.[27] In the 6-16B patterns the 3-12 instances are combined from the simultaneous major thirds and the last 16th notes in

---

[25] Each pattern comes with a simultaneous major third preceded by a 32nd note and followed by three 16th notes. For a description of other Vocabulary materials, see Forte (1982:160-3).

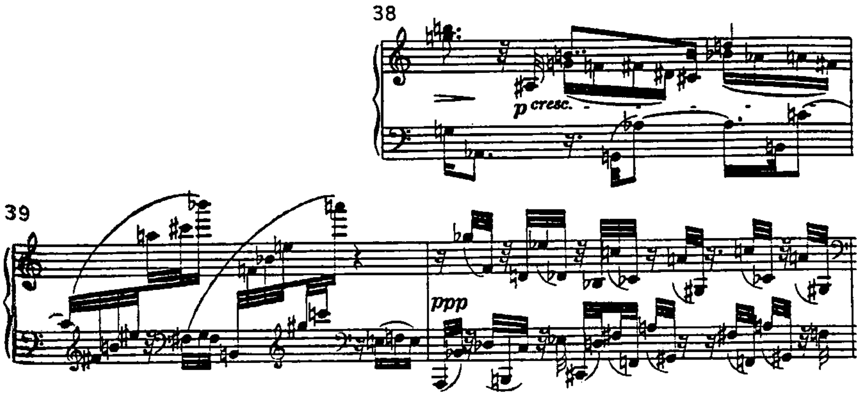
[26] This is the reason we characterized the 6-20 region as being of "horizontal orientation." Only selected pitches in the patterns qualify as segment members, the pitches between them being simply ignored.

[27] Authors disagree on the importance of SCs of cardinality 3, 3-12 among them, as structural entities in the movement. See Forte (1972:44) and (1982:136), Wittlich (1974:43,48), Perle (1981:14-5). We will consider the 3-12 instances as subformations of larger classes, not as independent structural elements.

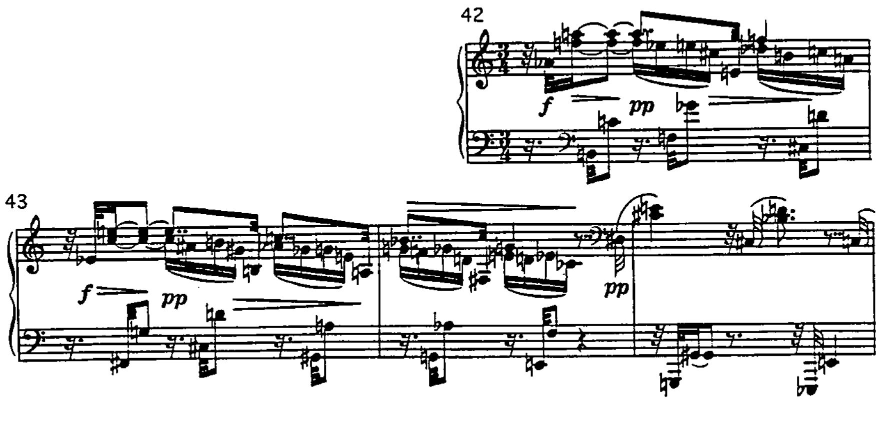
each pattern. The major thirds are often on the beat, the higher pitch being both the highest element in the pattern and the longest in duration. The highest pitch still sounds when the last 3-12 pitch is played.

Ex. 6.9.a gives the segmentation of the music in Ex. 6.9. Ex. 6.10.a, in turn, segments the Ex. 6.10 music. The 3-12 instances are also represented by chords.[28]

EXAMPLE 6.9: Mm. 38-40.

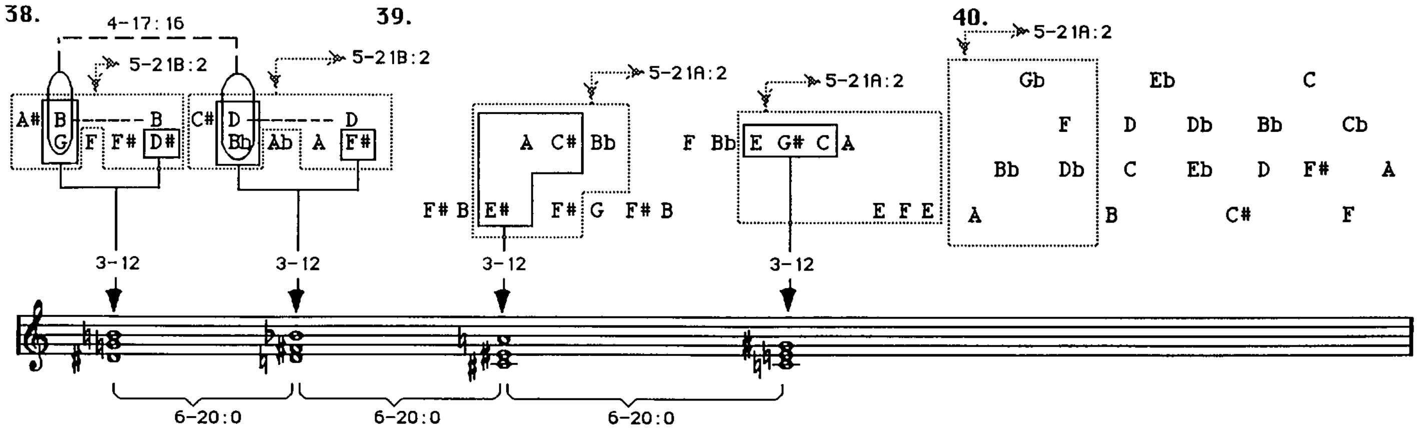


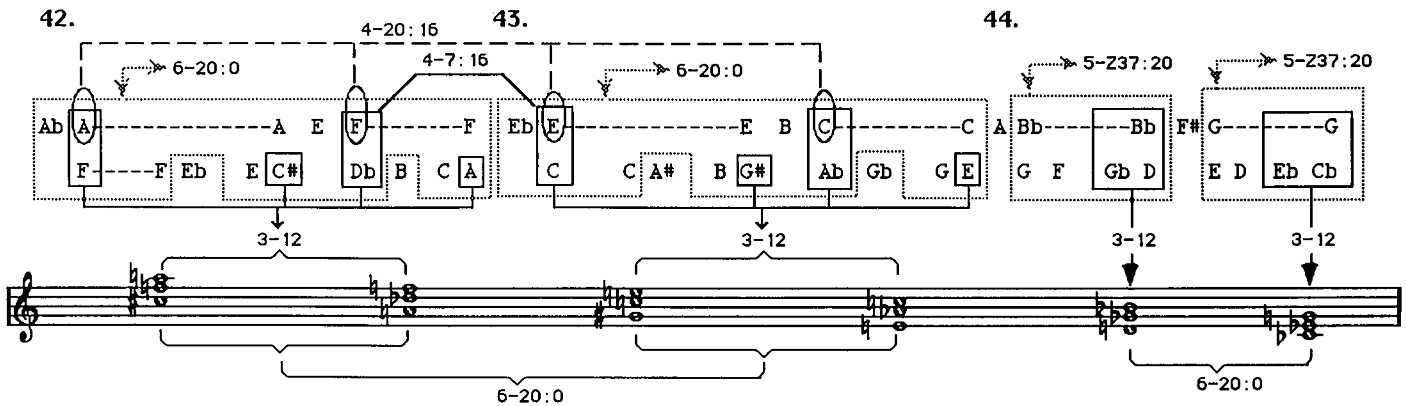
EXAMPLE 6.10: Mm. 42-45.



[28] The chords preserve the pc contents of the corresponding segments. Octave positions of the notes, however, may have been changed.

EXAMPLE 6.9.a: Mm. 38-40. Instances of SCs in the RECREL region 6-20. Value limit 20.





EXAMPLE 6.10.a: Mm. 42-44. Instances of SCs in the RECREL region 6-20. Value limit 20.

The 3-12 instances constitute successions of chromatically related pcsets. In Ex. 6.9.a, pairwise combinations of successive 3-12 instances produce three instances of the nexus SC 6-20.[29] In Ex. 6.10.a the harmonic rhythm is slower. The first and second 3-12 instances constitute the same pcset, {1,5,9}. The third and the fourth instances constitute also a single pcset, {0,4,8}. The succession of the four 3-12 instances produces only one instance of 6-20.

The 6-16B patterns are not the only formations containing the 3-12 instances. In m. 39 the instances E#-A-C# and E-G#-C are produced from successive pitches in the two arpeggiated patterns. (Ex. 6.9.a). M. 44, in turn, contains two patterns resembling the 6-16B patterns but representing another SC. In both patterns, an augmented triad can be combined from the higher note of the minor third (being the highest element registrally and longest durationally) and the two last sixteenth notes (Gb-D and Eb-Cb, respectively). The two 3-12 instances combine into another instance of 6-20. (Ex. 6.10.a). It should be noted that all 3-12 instances are embedded in instances of region 6-20 classes, 5-21A (2), 5-21B (2), 6-20 (0) and 5-Z37 (20). (Examples 6.9.a and 6.10.a).

In m. 38, the corresponding simultaneous major thirds in the two 6-16B patterns together produce an instance of SC 4-17 (16). (Ex. 6.9.a). The higher pitches in the four corresponding major thirds in mm. 42-43 constitute an instance of 4-20 (16). The second and third of the same major thirds, Db-F and C-E, together constitute an instance of 4-7 (16). Etc. (Ex. 6.10.a).

6.5.2.3.2 Nexus 8-28

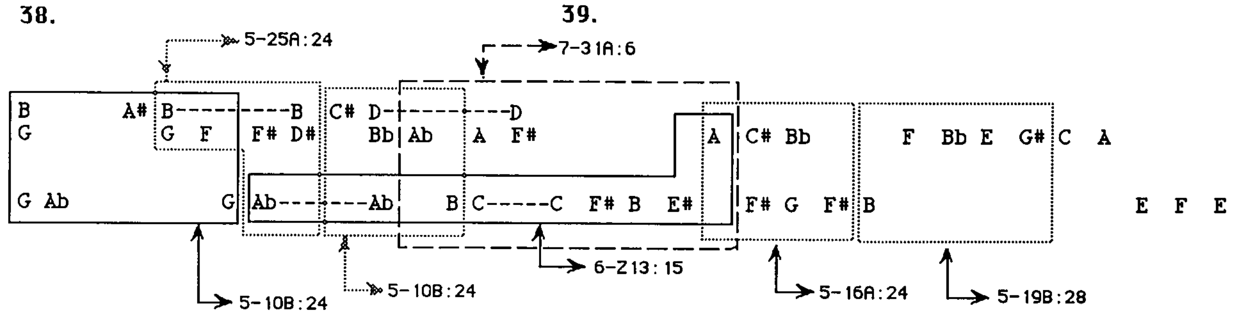
The value limit of the RECREL region 8-28 was set slightly higher than the limit 20 of the previous region. Some 5-pc classes had multiple instances in the music, and considering the relatively large size difference between them and the octad class nexus, their values - usually 24 - were deemed low enough.[30] All classes given below as 8-28 region members are also its subset-classes. Subset-classes with values over the limit will be shown as well, if they are considered to be of importance.

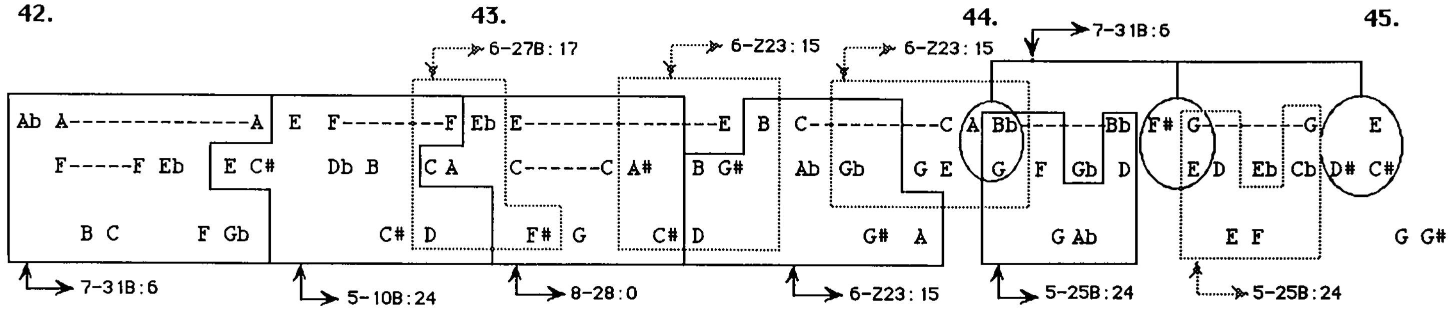
---

29 We could go a step further and speculate why the 3-12 instances are placed in just this sort of chromatic succession. When they are arranged in normal order, their corresponding elements constitute SC 4-1, not a Vocabulary class but one contained in the Most Prominent Sets in Wittlich (1974). Instances of this very same SC are extremely visible in the left-hand part of the previous measures, in mm. 35-38.

30 24 is the lowest value in the entire value group #5/#8. The average is 37.

EXAMPLE 6.11: Mm. 38-39. Instances of SCs in the RECREL region 8-28. Value limit 25.





EXAMPLE 6.12: M. 42 - beginning of m. 45. Instances of SCs in the RECREL region 8-28. Value limit 25.

With the exception of appearances of the two Vocabulary classes 6-Z13 (15) and 6-Z42 (25), the 8-28 region is a new element in the music, and contains many segments without clear textural profiles. The 6-20 region, on the other hand, has appeared already in two previous passages. Because of this, it would not be considered surprising if especially clear instances of classes in the new region were to be found in the music, consolidating its presence. Such instances are indeed to be found. Before these are examined in the next section, we will see how the 8-28 region strand is present in the same measures that were analysed above from the point of view of the 6-20 region. The measures segmented in Ex. 6.11 correspond approximately with those in 6.9, the measures in Ex. 6.12 with those in Ex. 6.10.

In Ex. 6.11, fragments of the two 6-16B statements are combined with coinciding and surrounding pitches to constitute instances of 8-28 region classes 5-10B (24), 5-25A (24), 6-Z13 (15) and 7-31A (6). The septad class and its inversional counterpart, 7-31B, are the sole 7-pc subset-classes of the nexus. The arpeggiated patterns in m. 39 contain two more pentad classes, 5-16A (24) and 5-19B (28). The value of the latter is above our limit, but it is nevertheless a subset-class of 8-28, obviously belonging to the region.

Ex. 6.12 suggests that the presence of the 8-28 region is stronger in mm. 42-44 than it was in the previous excerpt. SCs of cardinality 6 and larger predominate, resulting in lower RECREL values. This coincides with a small modification in two of the 6-16B patterns. The durations of the simultaneous major thirds are longer. The classes derived from the segments are 8-20 (0), 7-31B (6), several overlapping instances of 6-Z23 (15), 6-27B (17), and again some pentad classes with the value 24.

6.5.2.3.3 Interaction between the Two RECREL Regions

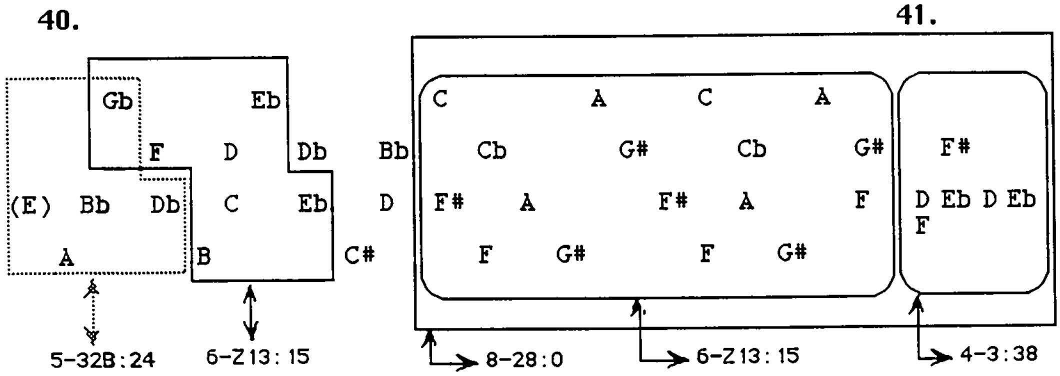
The developmental nature of the music in mm. 38-48 is evident, and, interestingly, the type of harmonic processing we examine with the two simultaneous RECREL regions seems to participate in producing this experience. The regions are not just passively present, but a dynamic interaction takes place between them.

Mm. 40-41

Our first example is from the beginning of the mm. 38-48 passage. In the 6-20 region, the first segments are more strongly profiled than those following in m. 39 and m. 40 (Ex. 6.9.a). The 8-28 region, in turn, was seen above as being not yet as strongly present as in mm. 42-44. Interestingly, these two notions, the 6-20 region "fading out" and the 8-28 region waiting to be consolidated, are followed by a momentary absence of the 6-20 region material and a forceful statement of set-class 8-28.

EXAMPLE 6.13: Mm. 40-41.

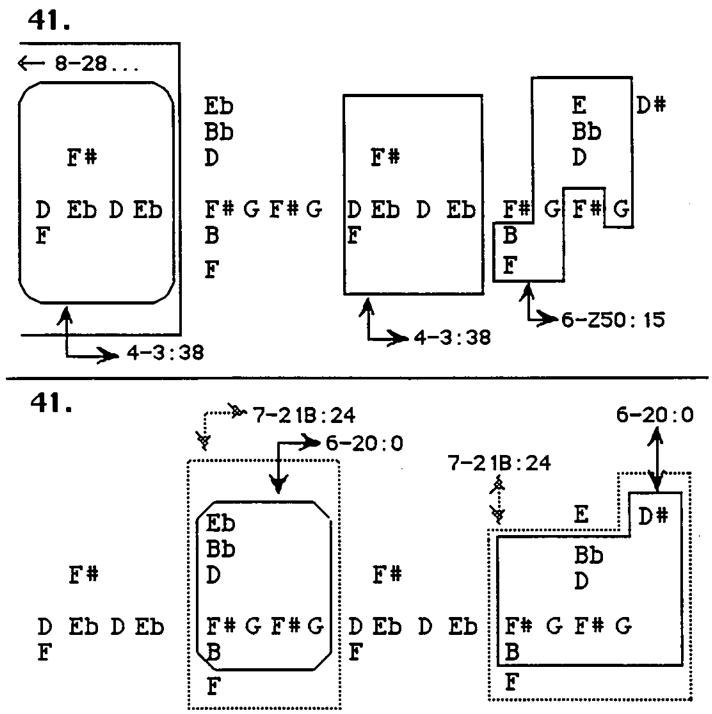
EXAMPLE 6.13.a: M. 40 - beginning of m. 41. Instances of set-classes in the RECREL region 8-28. Value limit 25.



Ex. 6.13.a gives the segmentation of m. 40 - beginning of m. 41 with 8-28 as the nexus. The "pointillistic" texture in m.40 contains first an instance of 5-32B (RECREL value 24 with 8-28), then an instance of the Vocabulary class 6-Z13 (15). Another instance of the latter SC follows, much emphasized on account of its length - 17 pitches - and the repetitions "freezing" the flow of music. In order to complete this formation into an instance of 8-28, two additional pitches, D and Eb, would be required. The pitches do follow immediately, becoming pointedly clear in the written tremolo at the beginning of m. 41. The pitch combination F-D-Eb-F# itself constitutes SC 4-3. Due to the large cardinality difference, the RECREL value between it and the nexus is rather high, 38.

A short "dialogue" follows in the next measure, 41. Elements from the two regions are arranged almost like different-coloured patches in a patchwork. This is illustrated in examples 6.13.b and 6.13.c. Both segment the same measure, but with different nexi, 8-28 and 6-20, respectively. The long 8-28 statement, out of which only the subset-class F-D-Eb-F# (SC 4-3) can be seen, is followed by an instance of 7-21B (RECREL value 24 with 6-20). Without the lowest F - a sort of "pedal point" pc as it is the lowest element in every chord in m. 41 - the same formation constitutes the nexus SC 6-20 itself. After an identical restatement of 4-3, the passage ends with a hybrid of the two regions: when the combination spanning the last beat of the measure is taken without F#, the lower pitch in the written tremolo, and D#, acting as a "resolution" to the "suspended" E, the combination constitutes 6-Z50 (15 with 8-28). When the same combination is taken without the E, the result is again 7-21B, containing the instance of 6-20 as its subset-class.

EXAMPLES 6.13.b and 6.13.c: Measure 41. 6.13.b (above line): Instances of SCs in the RECREL region 8-28. 6.13.c (below line): Instances of SCs in the RECREL region 6-20.

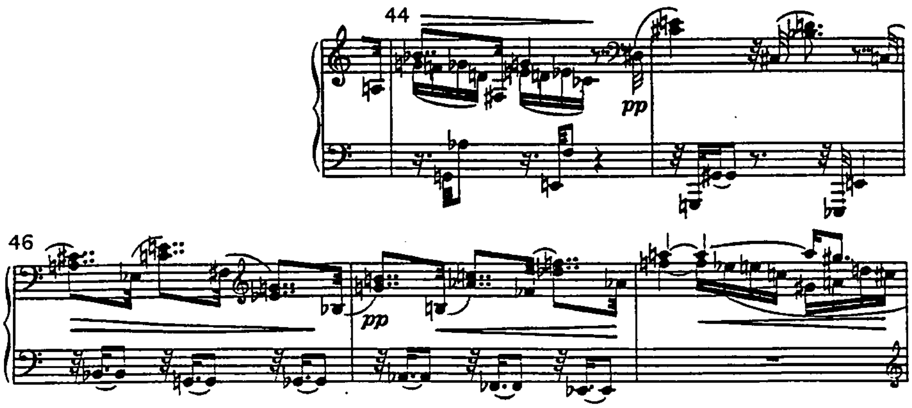


Mm. 44-48

Our second example of interaction between the RECREL regions is in mm. 44-48. The music in in Ex. 6.14. In outline the process is as follows: first, the 8-28 region assumes the function of its counterpart. It produces a horizontal 7-31B instance in mm. 44-45 in exactly the same manner as the 6-20 region has done up to that point, from corresponding pitches in recurring patterns. After this, the 6-20 region is again momentarily absent. A strong statement of SC 8-28 governs part of m. 45 and the whole m. 46. A rapid shift, almost like a modulation, from region 8-28 to region 6-20 takes place in mm. 46-47. The octatonic region fades out and the 6-20 region is consolidated.

The horizontal 8-28 region segment in m. 44, constituting 7-31B (6), is shown in Ex. 6.12 above. The measure contains two patterns texturally related to the preceding 6-16B patterns. Around the measure line 44-45 these are followed by a third, reduced pattern, within which only the simultaneous minor third and the preceding pitch remain. Together the three minor third-preceding pitch combinations produce the 7-31B instance.

EXAMPLE 6.14: Mm. 44-48.

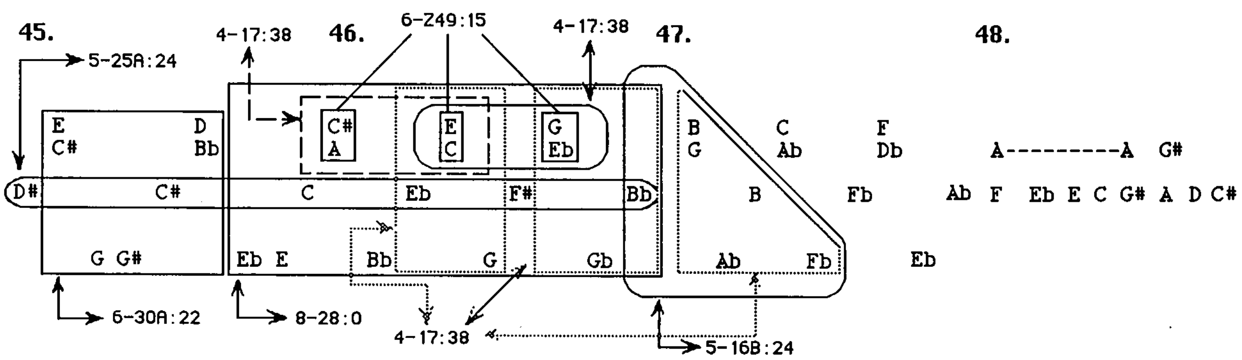


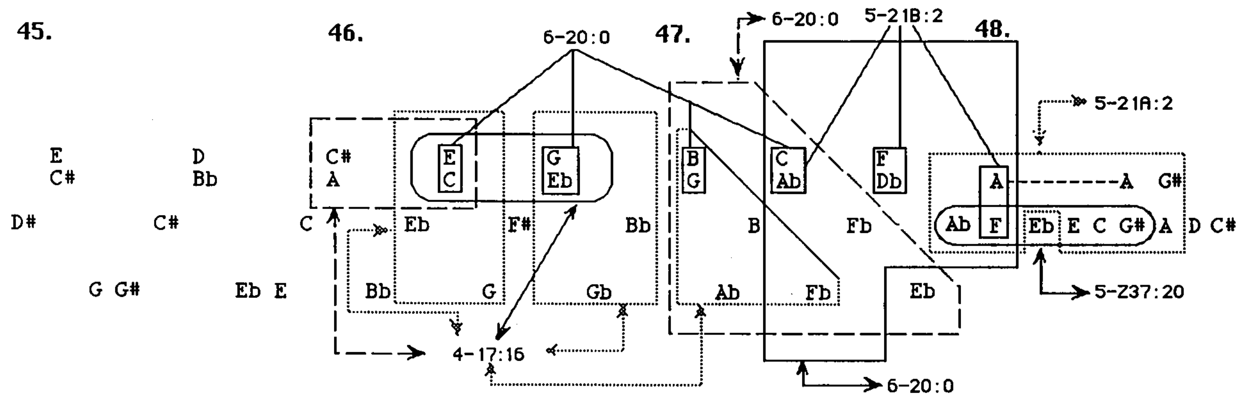
The 6-20 region is absent in m. 45, the 8-28 region providing two additional horizontal statements. (Ex. 6.14.a). The succession consisting of the six right-hand 32nd notes that precede the simultaneous thirds, D#-C#-C-Eb-F#-Bb, constitutes 5-25A (value 24 with 8-28). The three right-hand major thirds in m. 46 produce 6-Z49 (15). Among other 8-28 region instances are 6-30A (22) and the above-mentioned long instance of 8-28.[31]

The shift from the 8-28 region to the 6-20 region, in mm. 46-47, is shown with two different segmentations of the same measures (Ex. 6.14.a and Ex. 6.14.b). The octatonic class is the nexus in the former example.

31 Despite the fabric of 8-28 region instances, the most obvious element in the music is the strong left-hand statement of 6-Z10A.

EXAMPLE 6.14.a: M. 45 - beginning of m. 47. Instances of SCs in the RECREL region 8-28. Value limit 25.





EXAMPLE 6.14.b: Mm. 46-48. Instances of SCs in the RECREL region 6-20. Value limit 20.

The smoothness of the shift is related to the five instances of SC 4-17 in mm. 46-47, shown in both examples. 4-17 is the only class of cardinality four or larger included in both 8-28 and 6-20, and is being used as a link between the two regions.[32]

The two overlapping 4-17 instances produced by the successive right-hand major thirds in m. 46 are of special interest. When taken together, they constitute 6-Z49 (value 15 with 8-28. Ex.6.14.a). When taken with the major thirds B-G and C-Ab in m. 47, the latter instance constitutes 6-20. (Ex.6.14.b). The 6-20 instance, in turn, is linked to an instance of 5-21B (value 2 with 6-20) in mm. 47-48. The last of the 4-17 instances, in the beginning of m. 47, combines with the preceding Bb into an instance of 5-16B (value 24 with 8-28. Ex.6.14.a). With the simultaneous and following pitches C, Ab, Fb and Eb it produces again an instance of 6-20. (Ex.6.14.b). The 5-16B instance is the last element in the 8-28 region. The rest of the instances consolidate the 6-20 region at the end of the passage.

32 Due to the cardinality difference, the RECREL value between 8-28 and 4-17 is high, 38. Between 6-20 and 4-17 the value is 16.

■ CHAPTER 7
CONCLUSIONS

We believe that the main conclusion to be drawn from this study is the following: assessing SC similarity is a highly complicated task, calling for the participation of many more elements than actually participate in a number of existing similarity relations. This, no doubt, is reflected in some of the criticism in the theoretical literature. Commentators, even when spotting the potential benefits of some good relation, are sceptical that meaningful results can be obtained with the help of very simple comparison methods. We concur with this view.

In order to have a solid foundation for subsequent discussion, we established a set of criteria with which to evaluate various aspects of reliability and usefulness in similarity relations. When analysed with these criteria, almost all of the previously presented relations turned out to perform their tasks less than succesfully. Only one, Lewin's REL, was deemed acceptable.

Similarity relations not producing numeric values were seen to possess weak descriptive powers. The criteria with which they deem two SCs maximally or minimally similar were criticized as being inaccurate and concealing a gradation of intuitive similarity.

Among the actual similarity measures, two main approaches, based on pairing subset-class instances by one-to-one or by one-to-many correspondence, were identified. They were seen as having opposing advantages and disadvantages. One approach could not be deemed better than the other.

All measures comparing only one subset-class cardinality at a time were seen to produce at least some counterintuitive results, no matter how agreeable the comparison method itself seemed to be. Some measures also contained features that in our opinion distort their results and call their reliability into question. Besides the properties they claimed to observe, they ended up taking others into account as well. Furthermore, as we required that a measure should be able to discriminate between Z-related and inversionally related SCs, the total measures emerged as the most promising approach.

It was seen, however, that comparing subset-classes of all cardinalities is not unproblematic, either. In the values of some total measures, for example, certain subset-class cardinalities were seen to have a weak representation, in those of another measure a suspiciously strong one. Also, we argued that the intuitive starting point of measuring SC similarity - the acceptance of less-than-maximal dissimilarity between distinct SCs - must be applied to subset-classes as well. RECREL was then offered as an application of this principle and many others observed in connection with the previous measures.

Assessing the Validity of RECREL

The values RECREL produces were seen to correlate well with intuitive similarity assessments. On average RECREL deems small-cardinality SC pairs dissimilar, large-cardinality pairs similar. Transpositionally symmetric SCs participate often in the pairs with the very lowest values, but are on average distant from a majority of the classes. RECREL values belonging to inversionally related, Z-related, complement-related, inclusion-related (cardinalities n and n-1) and M-related pairs were obtained and analysed. Among I- and Z-pairs, even the most distant values suggested rather close similarity. On the other hand, among complement pairs, M-pairs and inclusion-related pairs of cardinalities n and n-1, values could fluctuate considerably. As far as Z-related and inversionally related SCs are concerned, all total measures produced rather similar results, challenging the notion of a constant degree of similarity between the classes.

An analysis of Schönberg's Opus 11, Number 1, suggested that RECREL can be of assistance in producing meaningful analytical results. The observations provided by it were in agreement with those of previous analyses concerning the principles with which certain structurally important SC materials are deployed in the music. It suggested that SC similarity is not a dominant aspect in the palindromic SC successions, with the exception of the areas around the centers of symmetry. The RECREL regions were used to examine aspects of the multi-dimensional harmonic structuring that previous studies have identified in atonal Schönberg. The measure helped in gathering SC materials whose harmonic characteristics were deemed similar enough to be analytically promising in this context. The way these SC materials were distributed and arranged, then, led to conclusions concerning harmonic coherence in the music. It was noted that several seemingly independent compositional processes could be active in the music at the same time. It was also found that the palindromic SC successions and the harmonic arrangements identified with the RECREL regions were not constantly present in the music, but in use only at times. They were seen as compositional devices that help to enrich and structure the movement, but not as principal elements from which the music stems.

Future Developments

It is evident that RECREL is only one application of the principles that were found relevant in assessing SC similarity. Future developments could, for example, begin by dividing the RECREL comparison procedure into two independent parts. The first part would consist of the internal similarity measure which is evaluated repeatedly during a comparison, a position presently occupied by $\%REL_n$. The second part, then, would be the process which identifies the unilaterally embedded subset-classes, arranges them into pairs for further comparisons, weights the values, updates the upper-level values with lower-level ones, etc. Instead of $\%REL_n$, the internal measure could be, for example, some development trying to combine the advantages of both one-to-one and one-to-many correspondence approaches, while avoiding their disadvantages. In the second part, the comparisons between unilaterally embedded subset-classes could be expanded. In the present version two such classes of cardinality n are compared only with respect to their subset-classes of cardinality n-1. The idea of a total measure could be applied here, too.

Combining similarity measures with other considerations constitute a whole category of potentially fruitful developments. For example, the measures could be used in designing other types of measure, those of pcset similarity. Pc collections are routinely compared, of course, but usually only with respect to their SC identities, not their distinct pcset identities. As writers note (Rahn 1989, Isaacson 1992:250-1), the actual compositional deployment of pcsets would greatly benefit from a "theory of instances."

Also the relation between SC similarity and aural similarity merits systematic exploration. Based on our own experience, we believe that large numbers of chord pairs derived from two SCs can offer meaningful reference points to purely abstract assessments of SC similarity. In each pair, the chords are to be as similar as possible with respect to registers, widths, pitch contents, movement between non-common pitches, and perhaps also spectral considerations, etc. One such application offered useful background information during the initial stages of the Schönberg analysis (section 6.2.1). Carefully designed and conducted listener group tests could be arranged in order to seek regularities in individual observations.

Experiences accumulated during chord pair comparisons could then be used while studying the relations between SC similarity and qualitative similarity (section 1.3). Let us envisage one possible test arrangement. Several groups of SC pairs would be selected so that, within each group, the pairs have a uniform value from a similarity measure but are associated with many different qualitative characteristics in the mind of the observer. The values would suggest increasing dissimilarity. Comparably presented chord pairs would again be derived from the SC pairs. The measured values, then, would offer exact points of reference when assessing (a) how chord pairs within a single group but with different qualitative characteristics relate to one another, and (b) how chord pairs from different groups but with similar qualitative characteristics relate to one another. Possible viewpoints for (a) could be, for example, whether the degree of aural similarity between the chord pairs seems to fluctuate, and if it did, whether some pattern explaining the fluctuations would emerge. With respect to (b) we could try to assess, for example, whether perceived discrimination of similarity seems to be coarser among dissonant chords or among consonant ones, among large chords or among small ones, etc. These results, in turn, could be of assistance in a task already mentioned in section 1.3, i.e., the development of weighting functions that assign degrees of dissonance for interval-classes and subset-classes.

Computer Applications

Because of the complexities of measuring SC similarity, theoretical considerations must be accompanied by a purely practical one, namely, the very act of *using* a measure. Calculating values by hand is arduous and sometimes, as when comparing SCs of large cardinalities with the total measures, practically impossible. Even consulting a matrix with hundreds of pre-calculated values is hopelessly counterintuitive for many. Obviously, the meaningfulness of using a similarity measure depends on the availability of a suitable computer application.

Use of a computer need not be confined to just assistance in calculating, sorting and filtering values, but it can be used also for more complicated, analysis-related tasks. A simple analytical application was developed for the purposes of the present study (section 6.5.1). The computer divided pitch successions into a large number of overlapping segments and compared these with RECREL. Despite certain awkward elements in the application, the results proved useful.

Ideally, similarity measures could become independent modules in analytical applications that are built around sophisticated score representation schemes. Among other modules could be measures of pcset and qualitative similarity, for example, as well as a wide range of powerful algorithms for pattern-matching, searching, sorting, filtering, etc. Such an application should also allow the user to define complex and detailed tasks which observe aspects of pitch organization in combination with any other musical parameters. Placed at the command of a skilled analyst, an application like this might produce interesting results supporting and complementing more traditional analytical practices.

■ GLOSSARY

The glossary summarizes the basic concepts associated with the RECREL similarity measure. The entries are given in an order that approximately follows the course of a RECREL comparison.

Recursive. Pertaining to a process that is inherently repetitive. The result of each repetition is usually dependent upon the result of the previous repetition. See entry "recursive routine."[1]

Recursive routine. In computing, a routine that may be used as a routine of itself, calling itself directly or being called by another routine, one that it itself has called.[2]

Difference Vectors. Given the comparison $\%REL_n(X,Y)$, a nonzero difference between corresponding components in the n-class %-vectors of X and Y indicates that some subset-class S is not represented with an equal share in X and Y. S may be mutually embedded in X and Y but with a different share, or it may be embedded in only one of the classes. In both alternatives, one of the SCs has a unilaterally embedded share of S in it. The nC%Vs reveal only indirectly which subset-classes have unilaterally embedded shares in the compared SCs, or how big the shares are. As this information will be needed during a RECREL comparison, it is given separately with the help of the difference vectors. To calculate these for X and Y from nC%V(X) and nC%V(Y), each pair of corresponding components $\{c_X,c_Y\}$ produces the differences $(c_X - c_Y)$ and $(c_Y - c_X)$. The non-negative $(c_X - c_Y)$ differences are entered in

[1] From Longley and Shain (1982), s.v. "recursive."

[2] From Longley and Shain (1982), s.v. "recursive routine."

the difference vector of X, the non-negative (cY - cX) differences in that of Y. Negative differences are replaced with zeros.

The two top rows of the example below give the 3-class %-vectors of SCs 5-1 and 5-23A, respectively. Below the indexes (row 3) are the difference vectors, giving the unilaterally embedded triad class shares in 5-1 and 5-23A. The SC names of the subset-classes are to be inferred from the indexes.

```
[30 20 20 10 10  0  0  0  0 10  0  0  0  0  0  0  0  0  0]
[ 0 10 10  0  0 10  0  0  0 10 10 20  0  0 20  0 10  0  0]

  1  2  2  3  3  4  4  5  5  6  7  7  8  8  9 10 11 11 12

[30 10 10 10 10  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
[ 0  0  0  0  0 10  0  0  0  0 10 20  0  0 20  0 10  0  0]

  1  2  2  3  3  4  4  5  5  6  7  7  8  8  9 10 11 11 12

[43 14 14 14 14  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
[ 0  0  0  0  0 14  0  0  0  0 14 29  0  0 29  0 14  0  0]
```

Scaled Difference Vectors. Vectors derived from difference vectors. Like difference vectors, except that their components are scaled so that their sum is always 100. They show how large a percentual representation each unilaterally embedded subset-class share has among all unilaterally embedded shares. To get the scaled version, each difference vector component is divided by the sum of all components and multiplied by 100. In the example above, the two lowest vectors are scaled difference vectors belonging to 5-1 and 5-23A.

Difference Group. A group of weight/subset-class pairs derived from a scaled difference vector (see **Weights**). In each pair, the first element is a nonzero scaled difference vector component, interpreted to be the weight of the second element, i.e., the subset-class to which the component belongs.

In the example of the *Difference Vector* entry above, the second lowest vector, a scaled difference vector belonging to SC 5-1, has five nonzero components. Together with the SC names to which their indexes refer, these components constitute the difference group {(43,3-1),(14,3-2A),(14,3-2B),(14,3-3A),(14,3-3B)}. Similarly, the lowest vector, belonging to 5-23A, produces the difference group {(14,3-4A),(14,3-7A),(29,3-7B),(29,3-9),(14,3-11A)}. Difference groups derived from a single comparison can never contain same SCs.

Cross-correlation Group. A combination of two difference groups. To obtain a cross-correlation group from difference groups A and B, each weight/subset-class element in A is paired with every element in B. If A and B contain n and m elements, respectively, the cross-correlation group will contain (n*m) elements. Several steps are needed before the weights and $\%REL_n$ values belonging to the SC pairs have been processed into a single value representing the whole cross-correlation group.

Weights. When the subset-class pairs in a cross-correlation group are compared with $\%REL_n$, there are two independent aspects to every comparison: the $\%REL_n$ value itself, and the weight the comparison has among all the comparisons. As a small component in a scaled difference vector indicates a modest standing for the subset-class having it, comparisons between such SCs can cover much smaller parts of the two unilaterally embedded subset-class materials than comparisons between subset-classes having large components. To eliminate the distortion an equal representation for each comparison would create, the components are interpreted to be the weights of the classes and used to determine the relative importance of each comparison.

Proportioned Weights. Combined weights derived from individual weights belonging to SCs in two difference groups. A proportioned weight gives the share a given SC pair has among all pairs in a cross-correlation group. If the two weights are w_1 and w_2, the proportioned weight of the pair is $(w_1 * w_2)/100$. The sum of all proportioned weights in a cross-correlation group is always 100.

Weighted Values. When both the $\%REL_n$ values and the proportioned weights have been calculated for all SC pairs in a cross-correlation group, each value v is weighted with its corresponding proportioned weight w_p. The weighted value $v_w = (v * w_p) /100$. A weighted value reflects both the degree of similarity between the SCs and the share the comparison has among all comparisons. The sum of all weighted values in a cross-correlation group gives the weighted arithmetic mean of the values.

Branches. Independent strands in a RECREL comparison. Each branch gets its own value, the final value being the average of all branch values. Given SCs X and Y of cardinality 6, for example, the highest branch is 5. The process to obtain the branch-5 value is begun by comparing the 5-class %-vectors of X and Y with $\%REL_5$, and deriving a cross-correlation group containing 5-pc subset-class pairs for further comparisons. Calculation of branch 4 starts with the comparison $\%REL_4(X,Y)$, branch 3 with the comparison $\%REL_3(X,Y)$, etc. The lowest branch to be calculated is 2.

Levels. $\%REL_n$ comparison categories within a single branch. Given SCs X and Y of cardinality 6 and branch 5, the first comparison, $\%REL_5(X,Y)$, takes place at the highest level, five. The cross-correlation group derived from 5C%V(X) and 5C%V(Y) contains pairs of pentad classes, $\%REL_4$ comparisons of which constitute branch 5, level 4. Each pentad class pair produces its own cross-correlation group. The $\%REL_3$ comparisons of the tetrad class pairs within these constitute branch 5, level 3, and so on. The lowest level is 2. After this comes branch 4, containing comparisons at levels 4, 3 and 2, etc.

Updating Values. Within branch n of the RECREL comparison between SCs X and Y, the level n contains only one comparison, $\%REL_n(X,Y)$. The level n-1 contains one group of comparisons, the level n-2 a group of groups of comparisons, etc, When all comparisons at levels from n to 2 have been completed, a single final value is determined for the whole branch. This is done by updating the value of each upper-level comparison with those of the lower-level comparisons derived from it. First, the sum of all weighted values is taken in each of the level-2 cross-correlation groups, resulting in their weighted arithmetic means. Each sum s of each cross-correlation group C updates the weighted value v_w of the level-3 SC pair P from which C was derived: $(s * v_w)/100$. The result is interpreted to be the new weighted value of P. P itself belongs to one of the level-3 cross-correlation groups. The sum of all (updated) weighted values is taken in each of these, and every group sum updates the value of the level-4 comparison from which the group was derived, etc. This is repeated until the weighted value sum of the only level-n-1 cross-correlation group has updated the sole level-n value $\%REL_n(X,Y)$. The result is the final value of the branch n.

■ APPENDIX
RECREL DEMO PROGRAM MANUAL

INTRODUCTION

With this program for the Macintosh computer, the user can explore all basic aspects of the RECREL similarity measure. The functions are divided into two categories. The first category contains functions for retrieving, sorting, filtering and analysing RECREL values. The second consists of functions illustrating the various internal aspects of a RECREL comparison: vectors, difference and cross-correlation groups, branch contents, branch results, etc.

GENERAL

To obtain a copy of the program

The program can be accessed through Internet and can be copied and distributed freely. The address is ftp.funet.fi and the directory pub/sci/incoming. The name of the file is recrel. The author can be contacted at the following addresses: castren@csc.fi (e-mail), or: Sibelius Academy, P.O. Box 86, 00251 Helsinki, Finland.

System requirements

A Macintosh computer with at least four megabytes of memory is required. Under System 7 or later, memory addressing must be set to 24 bits. The pre-set amount of memory reserved for the program is 6 megabytes, but can be easily adjusted in the Information window of the RECREL icon. In order to limit the size of the application the program was written with an older version of Common

Lisp. Because of this, the program may not work on 68040-based Macintosh models. If problems occur, try loading the program with the on-chip memory cache switched off.

An updated version of the program, being functionally identical from the point of view of the user but based on a newer Lisp, will replace the current version at a later date.

To run the program
Double-click the RECREL icon.

The windows
Once the program is loaded, two windows appear automatically. The first one is the **RECREL** window, where the functions are manipulated. The other one, the Listener window, is where the results are shown.

The **RECREL** window contains **buttons, editable text boxes** and **static texts.** The horizontal line indicates the division of the functions into the two categories described above.

Each of the 17 buttons - the boxes with rounded edges and text in them - controls one function. A function is evaluated by clicking a button once.

The editable text boxes - the 30 sharp-edged boxes with SC names or numbers in them - contain arguments for the functions. The number of arguments used by the functions varies from one to six. They are to be found to the right of the functions using them, with the exception of the three arguments preceded by the text "arguments for all functions below." The pre-set SC names and numbers in the editable text boxes have no special significance. They constitute a precaution only, since evaluating a function with empty arguments would cause an error message. To change an argument, place the cursor to the right of the current one, click and drag leftwards while holding the button down, then enter a new SC name or number.

The static texts below the editable text boxes indicate the specific arguments which the boxes control.

Changing a font
When vectors wider than will fit on a small screen must be examined, a few alternatives are available: using a smaller font in the **Listener** window, making **Listener** broader than the screen, or both. To change the font, click into **Listener** to make it the active window. Choose **Font** from the **Edit** menu and select the **Courier** font. The output can be condensed even further by choosing

Font Style in **Edit** and selecting **Condense**. If the result is already too condensed, activate **Plain** in **Font Style**. Only **Courier** or **Monaco** fonts are recommended as the characters in the other fonts are not evenly spaced.

Adjusting the Listener screen size

Activate **Listener** by clicking into it. Point the handle bar (the area where the word **Listener** is written), click and drag leftwards while holding the button down. Part of **Listener** is now out of sight. Point to the Size box (the small square at the lower right-hand corner of **Listener**), click and drag rightwards while holding the button down. The window becomes larger. Use the horizontal scrollbar at the bottom of **Listener** to scroll left and right in the window.

Highlighting parts of the result lists

Some of the functions, especially Branch Contents 1 and 2, return typical Lisp results containing nested lists. Highlighting different parts of a result list is an excellent way of studying its composition. Seek the group of closing parentheses,), at the very end of a result list and place the cursor immediately to the right of the leftmost one. Double-clicking highlights the innermost list. Move the cursor one step to the right, double-click, and a list of lists is highlighted, and so on.

Printing on paper

The results can be printed on paper. Choose **New** in the **File** menu. A new window appears. Go to the **Listener** window, copy the material you want and paste it to the new window. Choose **Print** from the **File** menu.

"Garbage collections"

From time to time Lisp will interrupt the normal execution of the program for about 2-14 seconds. The length of time depends on the Macintosh model. During these interruptions, or *garbage collections*, used memory cells are "recycled." When a garbage collection takes place, the cursor is transformed into the letter combination GC.

Some examples below may have been slightly edited for reasons of clarity.

THE FUNCTIONS

1 MANIPULATING THE RECREL VALUES

RECREL Value

Description: Returns RECREL values belonging to pairs of SCs.

Usage:[1] RECREL-Value *SC1 SC2*

Details: The values are stored in a precalculated table. Retrieving them is fast. *SC1* and *SC2* may be any two SCs between cardinalities 2 and 12. The order between them is free.

Example: (RECREL-Value 3-1 3-2A) ***returns*** `RECREL(3-1,3-2A) = 33`

SC Name Help

Description: Returns the SC names in a cardinality-class.

Usage: SC-Name-Help *cardinality-class*

Details: Before a given RECREL demo program function is evaluated, the arguments are sent to a function named Argument-Check. Among other things this function examines that SC names are written correctly. If a mistake is detected or a SC name given without the extra labels A or B used in this study, an error message is generated and the evaluation of the original function prevented. When in doubt about the proper way of writing SC names, use SC-Name-Help. *cardinality-class* may vary between 2 and 12.

Example: (SC-Name-Help 3) ***returns***

```
Cardinality-Class 3:

(3-1 3-2A 3-2B 3-3A 3-3B 3-4A 3-4B 3-5A 3-5B 3-6 3-7A 3-7B 3-8A 3-8B 3-9
3-10 3-11A 3-11B 3-12)
```

---

[1] In the **Usage** and **Example** entries, each function name *Function Name* will be written as a single character string *Function-Name*, in order to make the examples resemble actual Lisp function evaluations more closely.

Ind Val Group SC/#n-#m

Description: Forms an individual value group and returns information describing its value distribution.

Usage: Ind-Val-Group-SC/#n-#m *SC n m*

Details: As the individual value group *SC/#n-#m* may contain hundreds of values, the function does not return the group itself. The result is a table containing details of the group. The entries are the following: lowest, average and highest values; number of all values and distinct values; percentiles, giving the number of values below 10th percentile, 10th, 25th, 50th (the median), 75th and 90th percentiles and the number of values above the 90th percentile.

SC may be any SC between cardinalities 2 and 12. $n - m$ may be any range of cardinality-classes where (1) $n \le m$, (2) $2 \le n,m \le 12$. Whenever $n \le \#SC \le m$, the value of the pair *{SC,SC}* is omitted from the value group.

Example: (Ind-Val-Group-SC/#n-#m 3-1 3 3) ***returns***

```
Individual Value Group:  3-1/#3-#3

Lowest Value:          33    Average Value:  71   Highest Value: 100

Number of Values:      18    Number of Distinct Values:   3

Percentiles:
  # of Vals < 10th:   2
  10th:               43.2
  25th:               67.0
  Median:             67.0
  75th:               67.0
  90th:               100.0
  # of Vals > 90th:   0
```

num Closest in SC/#n-#m

Description: Returns a required number of lowest values in an individual value group.

Usage: num-Closest-in-SC/#n-#m *SC num n m*

Details: First, the individual value group *SC/#n-#m* is formed. The values are then sorted in ascending order and *num* first values returned as the result. If *num* exceeds the number of all values in *SC/#n-#m,* the entire sorted value group is returned.

SC may be any SC between cardinalities 2 and 12. *n - m* may be any range of cardinality-classes where $n \leq m$ and $2 \leq n,m \leq 12$. Whenever $n \leq \#SC \leq m$, the value of the pair *{SC,SC}* is omitted from the value group. Each sublist in the result contains *SC,* another SC and the RECREL value between the two.[2]

Example: (num-Closest-in-SC/#n-#m 3-1 10 3 4) ***returns***

```
The 10 SC pairs with the lowest values in the
Individual Value Group 3-1/#3-#4:

((3-1 4-1 17) (3-1 3-2A 33) (3-1 3-2B 33) (3-1 4-2A 38) (3-1 4-2B 38)
 (3-1 4-3 50) (3-1 4-4A 50) (3-1 4-4B 50) (3-1 4-5A 50) (3-1 4-5B 50))
```

Lim1 -> Lim2 in SC/#n-#m

Description: Returns all values between required lower and upper limits in an individual value group.

Usage: Lim1->Lim2-in-SC/#n-#m *SC n m limit1 limit2*

Details: When the individual value group *SC/#n-#m* is formed, only the values between *limit1* and *limit2,* inclusive, are retained. These are sorted in ascending order and returned as the result. *limit1 - limit2* may be any range of values so that $limit1 \leq limit2$ and $0 \leq limit1,limit2 \leq 100$. When *limit1* = 0 and *limit2* = 100, the entire sorted value group *SC/#n-#m* is returned.

SC may be any SC between cardinalities 2 and 12. *n - m* may be any range of cardinality-classes where $n \leq m$ and $2 \leq n,m \leq 12$. Whenever $n \leq \#SC \leq m$, the value of the pair *{SC,SC}* is omitted from the value group. Each sublist in the result contains *SC,* another SC and the RECREL value between the two.

Example: (Lim1->Lim2-in-SC/#n-#m 3-1 3 4 0 40) ***returns***

```
The SC Pairs with values between 0 and 40 in the
Individual Value Group 3-1/#3-#4:

((3-1 4-1 17) (3-1 3-2B 33) (3-1 3-2A 33) (3-1 4-2B 38) (3-1 4-2A 38))
```

---

[2] As the sublists contain both SC pairs and values, the results are in fact combinations of (sorted and shortened) comparison groups and value groups. For the sake of simplicity we will refer to them here only as value groups. Similar result list composition will be used with a number of other functions as well.

Val Grp #n1-#m1/#n2-#m2

Description: Forms a value group and returns information describing its value distribution.

Usage: Val-Grp-#n1-#m1/#n2-#m2 *n1 m1 n2 m2*

Details: As the value group *#n1-#m1/#n2-#m2* may contain thousands of values, the function does not return the group itself. The result is a table containing details of the group. The entries are as follows: lowest, average and highest values; number of all values and distinct values; percentiles, giving the number of values below 10th percentile, 10th, 25th, 50th (the median), 75th and 90th percentiles and the number of values above the 90th percentile.

To compile *#n1-#m1/#n2-#m2*, each SC in the range of cardinality-classes *n1 - m1* is compared to every SC in the range of cardinality-classes *n2 - m2*. Whenever the two ranges intersect, the values of all SC pairs {X,X} are omitted from the value group, a given pair {X,Y} = {Y,X} providing only one value.

n1 - m1 and *n2 - m2* may be any two ranges of cardinality-classes where (1) $n1 \leq m1$, (2) $n2 \leq m2$, (3) $2 \leq n1,m1,n2,m2 \leq 12$.

Important: Caution is to be exercised when selecting the two ranges of cardinality-classes. Two broad ranges may result in a value group containing tens of thousands of values, taking a substantial amount of calculation time. It is recommended that the function is first tested with the smallest value groups.

Example: (Val-Grp-#n1-#m1/#n2-#m2 3 3 4 4) ***returns***

```
Value Group:  #3-#3/#4-#4

Lowest Value:          0    Average Value:  56  Highest Value: 100

Number of Values:    817    Number of Distinct Values:  13

Percentiles:
  # of Vals < 10th:  76
  10th:              38.0
  25th:              50.0
  Median:            54.0
  75th:              67.0
  90th:              83.0
  # of Vals > 90th:  14
```

Lim1 -> Lim2 in Val Grp

Description: Returns all values between required lower and upper limits in a value group.

Usage: Lim1->Lim2-in-Val-Grp *n1 m1 n2 m2 limit1 limit2*

Details: When the value group *#n1-#m1/#n2-#m2* is formed, only the values between *limit1* and *limit2*, inclusive, are retained. These are sorted in ascending order and returned as the result.

limit1 - limit2 may be any range of values where (1) $limit1 \leq limit2$, (2) $0 \leq limit1, limit2 \leq 100$. When $limit1 = 0$ and $limit2 = 100$, the entire sorted value group *#n1-#m1/#n2-#m2* is returned.

n1 - m1 and *n2 - m2* may be any two ranges of cardinality-classes where (1) $n1 \leq m1$, (2) $n2 \leq m2$, (3) $2 \leq n1, m1, n2, m2 \leq 12$.

Important: Caution is to be exercised when selecting the two ranges of cardinality-classes. It is recommended that the function is first tested with small value groups and narrow distances between *limit1* and *limit2.*

Example: (Lim1->Lim2-in-Val-Grp 3 3 4 4 0 20) ***returns***

```
The SC pairs with values between 0 and 20 in the
Value Group #3-#3/#4-#4:

((3-10 4-28 0)    (3-8B 4-25 0)     (3-8A 4-25 0)     (3-5B 4-9 0)
 (3-5A 4-9 0)     (3-11B 4-26 17)   (3-11B 4-20 17)   (3-11B 4-17 17)
 (3-11A 4-26 17)  (3-2A 4-3 17)     (3-11A 4-20 17)   (3-11A 4-17 17)
 (3-9 4-23 17)    (3-8B 4-21 17)    (3-8A 4-21 17)    (3-7B 4-26 17)
 (3-7B 4-23 17)   (3-7B 4-10 17)    (3-7A 4-26 17)    (3-7A 4-23 17)
 (3-7A 4-10 17)   (3-6 4-21 17)     (3-5B 4-8 17)     (3-5A 4-8 17)
 (3-4B 4-20 17)   (3-4B 4-8 17)     (3-4B 4-7 17)     (3-4A 4-20 17)
 (3-4A 4-8 17)    (3-4A 4-7 17)     (3-3B 4-17 17)    (3-3B 4-7 17)
 (3-3B 4-3 17)    (3-3A 4-17 17)    (3-3A 4-7 17)     (3-3A 4-3 17)
 (3-2B 4-10 17)   (3-2B 4-3 17)     (3-2B 4-1 17)     (3-2A 4-10 17)
 (3-2A 4-1 17)    (3-1 4-1 17))
```

2 FUNCTIONS ILLUSTRATING ASPECTS OF A RECREL COMPARISON

Important: Unlike the functions above, those below do not retrieve their results from precalculated tables, but calculate them separately each time. In order to prevent the RECREL demo program from becoming too large, as well as to ensure that unrealistically demanding comparisons are not attempted, some restrictions were placed upon the arguments which the functions will accept. No SCs of cardinalities larger than seven may be entered as arguments. For n-class vectors, n-class %-vectors, $\%REL_n$ comparisons, etc., the highest accepted n is five. n may not exceed the cardinalities of the SCs. Incorrect arguments will cause an error message.

All functions below perform calculations with full accuracy, but round their final results either to the nearest integers or to one decimal place. In the examples of Chapter four, however, already intermediate results were rounded to one decimal place. When comparing Chapter four results to those returned by the demo program, slight differences may therefore occur.

As the components in %-vectors, the weights in difference groups, etc., are shown rounded, their sum may not always be exactly 100. A vector component 100 is replaced with C.

nCV(SC) and nC%V(SC)

Description: Returns n-class vectors and n-class %-vectors.

Usage: nCV(SC)-and-nC%V(SC) *SC n*

Details: The result gives *nCV(SC)* first, then *nC%V(SC)* and below these, a row of indexes.
Limits for the arguments: (1) $2 \leq SC \leq 7$, (2) $2 \leq n \leq 5$, (3) $n \leq SC$.

Example: (nCV(SC)-and-nC%V(SC) 5-1 3) ***returns***

```
Top: 3CV(5-1)   Middle: 3C%V(5-1)   Bottom: Indexes

  3  2  2  1  1  0  0  0  0  1  0  0  0  0  0  0  0  0  0
 30 20 20 10 10  0  0  0  0 10  0  0  0  0  0  0  0  0  0
  1  2  2  3  3  4  4  5  5  6  7  7  8  8  9 10 11 11 12
```

$\%REL_n$(SC1,SC2)

Description: Returns $\%REL_n$ values.

Usage: $\%REL_n$ *SC1 SC2 n*

Details: Returns the (rounded) value of the comparison $\%REL_n(SC1,SC2)$. Limits for the arguments: (1) $2 \leq \#SC1,\#SC2 \leq 7$, (2) $2 \leq n \leq 5$, (3) $n \leq \#SC1,\#SC2$.

Example: ($\%REL_n$ 5-1 5-35 3) ***returns*** `%REL3(5-1,5-35) = 90`

nC%V & Diff Vec Pairs

Description: Returns pairs of n-class %-vectors, difference vectors and scaled difference vectors.

Usage: nC%V-&-Diff-Vec-Pairs *SC1 SC2 n*

Details: The *n*-class %-vectors of *SC1* and *SC2* are given first, then a row of indexes, the difference vectors of *SC1* and *SC2*, respectively, another row of indexes and, finally, the scaled difference vectors of *SC1* and *SC2*, respectively. The value of the comparison $\%REL_n(SC1,SC2)$ is given above the vector pairs. Limits for the arguments: (1) $2 \leq \#SC1,\#SC2 \leq 7$, (2) $2 \leq n \leq 5$, (3) $n \leq \#SC1,\#SC2$.

Example: (nC%V-&-Diff-Vec-Pairs 5-1 5-35 3) ***returns***

```
%REL3(5-1,5-35) = 90
 Top pair: 3C%V(5-1) and 3C%V(5-35)
 Middle pair: Difference Vectors
 Bottom pair: Scaled Difference Vectors
 3rd and 6th rows: Indexes

 30 20 20 10 10  0  0  0  0 10  0  0  0  0  0  0  0  0  0
  0  0  0  0  0  0  0  0  0 10 20 20  0  0 30  0 10 10  0

  1  2  2  3  3  4  4  5  5  6  7  7  8  8  9 10 11 11 12

 30 20 20 10 10  0  0  0  0  0  0  0  0  0  0  0  0  0  0
  0  0  0  0  0  0  0  0  0  0 20 20  0  0 30  0 10 10  0

  1  2  2  3  3  4  4  5  5  6  7  7  8  8  9 10 11 11 12

 33 22 22 11 11  0  0  0  0  0  0  0  0  0  0  0  0  0  0
  0  0  0  0  0  0  0  0  0  0 22 22  0  0 33  0 11 11  0
```

Difference Groups

Description: Returns difference groups.

Usage: Difference-Groups *SC1 SC2 n*

Details: The difference groups are derived from the *n*-class %-vectors of *SC1* and *SC2*, first by determining the difference vectors and scaled difference vectors, then collecting the nonzero components (weights) and their indexes (subset-class names) from the latter.

As difference groups are not derived from level-two comparisons, the limits for the arguments are as follows: (1) $3 \leq \#SC1,\#SC2 \leq 7$, (2) $3 \leq n \leq 5$, (3) $n \leq \#SC1,\#SC2$.

Besides the difference groups, the result gives also the value of the comparison $\%REL_n(SC1,SC2)$.

Example: (Difference-Groups 5-1 5-35 3) ***returns***

```
The two Difference Groups derived from 3C%V(5-1) and 3C%V(5-35).
%REL3(5-1,5-35) = 90. Each sublist contains a weight and a SC name

  Difference Group of 5-1:
((33 3-1) (22 3-2A) (22 3-2B) (11 3-3A) (11 3-3B))

  Difference Group of 5-35:
((22 3-7A) (22 3-7B) (33 3-9) (11 3-11A) (11 3-11B))
```

Cross-Corr Group 1

Description: Returns cross-correlation groups containing weights and SC names.

Usage: Cross-Corr-Group-1 *SC1 SC2 n*

Details: The cross-correlation group is derived from the *n*-class %-vectors of *SC1* and *SC2*. First, the difference vectors, scaled difference vectors and difference groups are derived from the two nC%Vs. Then, each SC in one difference group is combined with every SC in the other. The result is a level-*n-1* cross-correlation group.

Cross-correlation groups are not derived from level-two comparisons. Accordingly, the limits for *SC1, SC2* and *n* are the following: (1) $3 \le \#SC1,\#SC2 \le 7$, (2) $3 \le n \le 5$, (3) $n \le \#SC1,\#SC2$.

Each sublist in the result refers to one $\%REL_{n-1}$ comparison, containing first the weights of the two subset-classes belonging to *SC1* and *SC2*, respectively, then the subset-classes themselves. The result gives also the value of the comparison $\%REL_n(SC1,SC2)$, as well as the size of the cross-correlation group.

Example:[3] (Cross-Corr-Group-1 5-1 5-35 3) ***returns***

```
A level-2 Cross-Correlation Group derived from 3C%V(5-1) and 3C%V(5-35).
%REL3(5-1,5-35) = 90.
In each sublist, the four entries are:
(1) the weight of the subset-class belonging to 5-1,
(2) the weight of the subset-class belonging to 5-35,
(3) the subset-class belonging to 5-1,
(4) the subset-class belonging to 5-35.
Number of sublists: 25.

(33 22 3-1  3-7A) (33 22 3-1  3-7B) (33 33 3-1  3-9) (33 11 3-1  3-11A)(33 11 3-1  3-11B)
(22 22 3-2A 3-7A) (22 22 3-2A 3-7B) (22 33 3-2A 3-9) (22 11 3-2A 3-11A)(22 11 3-2A 3-11B)
(22 22 3-2B 3-7A) (22 22 3-2B 3-7B) (22 33 3-2B 3-9) (22 11 3-2B 3-11A)(22 11 3-2B 3-11B)
(11 22 3-3A 3-7A) (11 22 3-3A 3-7B) (11 33 3-3A 3-9) (11 11 3-3A 3-11A)(11 11 3-3A 3-11B)
(11 22 3-3B 3-7A) (11 22 3-3B 3-7B) (11 33 3-3B 3-9) (11 11 3-3B 3-11A)(11 11 3-3B 3-11B)
```

3 Result list printed with a smaller font to illustrate the cross-correlation group composition better.

Cross-Corr Group 2

Description: Returns cross-correlation groups containing proportioned weights, SC names and %REL_n values.

Usage: Cross-Corr-Group-2 *SC1 SC2 n*

Details: The cross-correlation group is derived from the n-class %-vectors of *SC1* and *SC2*. First, the difference vectors, scaled difference vectors and difference groups are derived from the two nC%Vs. Then, each SC in one difference group is combined with every SC in the other. The result is a level-n-1 cross-correlation group.

Cross-correlation groups are not derived from level-two comparisons. Accordingly, the limits for *SC1, SC2* and n are as follows: (1) $3 \le$ *#SC1,#SC2* ≤ 7, (2) $3 \le n \le 5$, (3) $n \le$ *#SC1,#SC2*.

Each sublist in the result refers to one %REL_{n-1} comparison and contains first the proportioned weight of the two subset-classes belonging to *SC1* and *SC2*, respectively, then the subset-classes themselves, and as the last entry the %REL_{n-1} value between the subset-classes. The result gives also the value of the comparison %REL_n*(SC1,SC2)* and the size of the cross-correlation group.

Example:[4] (Cross-Corr-Group-2 5-1 5-35 3) ***returns***

```
A level-2 Cross-Correlation Group derived from 3C%V(5-1) and 3C%V(5-35).
%REL3(5-1,5-35) = 90.
In each sublist, the four entries are:
(1) the proportioned weight of the subset-class pair,
(2) the subset-class belonging to 5-1,
(3) the subset-class belonging to 5-35,
(4) the %REL2 value of the subset-class pair.
Number of sublists: 25.

((7 3-1  3-7A 67) (7 3-1  3-7B 67) (11 3-1 3-9 67)  (4 3-1  3-11A 100) (4 3-1 3-11B 100)
 (5 3-2A 3-7A 33) (5 3-2A 3-7B 33) (7 3-2A 3-9 67)  (2 3-2A 3-11A 67)  (2 3-2A 3-11B 67)
 (5 3-2B 3-7A 33) (5 3-2B 3-7B 33) (7 3-2B 3-9 67)  (2 3-2B 3-11A 67)  (2 3-2B 3-11B 67)
 (2 3-3A 3-7A 67) (2 3-3A 3-7B 67) (4 3-3A 3-9 100) (1 3-3A 3-11A 33)  (1 3-3A 3-11B 33)
 (2 3-3B 3-7A 67) (2 3-3B 3-7B 67) (4 3-3B 3-9 100) (1 3-3B 3-11A 33)  (1 3-3B 3-11B 33))
```

[4] Result list printed with a smaller font to illustrate the cross-correlation group composition better.

Cross-Corr Group 3

Description: Returns cross-correlation groups containing SC names and weighted %REL_n values.

Usage: Cross-Corr-Group-3 *SC1 SC2 n*

Details: The cross-correlation group is derived from the *n*-class %-vectors of *SC1* and *SC2*. First, the difference vectors, scaled difference vectors and difference groups are derived from the two nC%Vs. Then, each SC in one difference group is combined with every SC in the other. The result is a level-*n-1* cross-correlation group.

Cross-correlation groups are not derived from level-two comparisons. Accordingly, the limits for *SC1, SC2* and *n* are as follows: (1) $3 \leq \#SC1, \#SC2 \leq 7$, (2) $3 \leq n \leq 5$, (3) $n \leq \#SC1, \#SC2$.

Each sublist in the result refers to one %REL_{n-1} comparison, containing two subset-classes belonging to *SC1* and *SC2*, respectively, and the weighted %REL_{n-1} value of the subset-class pair. The result gives also the value of the comparison %$REL_n(SC1,SC2)$ and the size of the cross-correlation group.

Example: (Cross-Corr-Group-3 5-1 5-35 3) ***returns***

```
A level-2 Cross-Correlation Group derived from 3C%V(5-1) and 3C%V(5-35).
%REL3(5-1,5-35) = 90.
In each sublist, the three entries are:
(1) the subset-class belonging to 5-1,
(2) the subset-class belonging to 5-35,
(3) the weighted %REL2 value of the subset-class pair.
Number of sublists: 25.

((3-1  3-7A 5) (3-1  3-7B 5) (3-1  3-9 7) (3-1  3-11A 4) (3-1  3-11B 4)
 (3-2A 3-7A 2) (3-2A 3-7B 2) (3-2A 3-9 5) (3-2A 3-11A 2) (3-2A 3-11B 2)
 (3-2B 3-7A 2) (3-2B 3-7B 2) (3-2B 3-9 5) (3-2B 3-11A 2) (3-2B 3-11B 2)
 (3-3A 3-7A 2) (3-3A 3-7B 2) (3-3A 3-9 4) (3-3A 3-11A 0) (3-3A 3-11B 0)
 (3-3B 3-7A 2) (3-3B 3-7B 2) (3-3B 3-9 4) (3-3B 3-11A 0) (3-3B 3-11B 0))
```

Branch Contents 1

Description: Displays the level composition of a branch as well as the contents of the individual levels, returning proportioned weights, SC names and $\%REL_n$ values.

Usage: Branch-Contents-1 *SC1 SC2 n*

Details: The function gathers all $\%REL_n$ comparisons in branch *n* of the RECREL comparison between *SC1* and *SC2*. The level composition can be inferred from the indentation of the result. The sole level-*n* comparison is the leftmost one at the top. The comparisons at each of the lower levels *n-1*, *n-2*,...2 are laid out stepwise to the right. Several level-two comparisons may be placed on a single row.

Each sublist in the result contains a proportioned weight, two SCs and a $\%REL_n$ value. The proportioned weights and values are rounded to one decimal place.

Limits for the arguments: (1) $2 \leq \#SC1,\#SC2 \leq 7$, (2) $2 \leq n \leq 5$, (3) $n \leq \#SC1,\#SC2$.

Example:[5] (Branch-Contents-1 5-1 5-35 3) ***returns***

```
Branch 3 of the RECREL comparison between SCs 5-1 and 5-35.
The entries in each sublist: a proportioned weight, two SCs and a %RELn
value.

(((100.0 5-1 5-35 90.0)
  ((7.4 3-1 3-7A 66.7)    (7.4 3-1 3-7B 66.7)    (11.1 3-1 3-9 66.7)
   (3.7 3-1 3-11A 100.0) (3.7 3-1 3-11B 100.0) (4.9 3-2A 3-7A 33.3)
   (4.9 3-2A 3-7B 33.3)  (7.4 3-2A 3-9 66.7)   (2.5 3-2A 3-11A 66.7)
   (2.5 3-2A 3-11B 66.7) (4.9 3-2B 3-7A 33.3)  (4.9 3-2B 3-7B 33.3)
   (7.4 3-2B 3-9 66.7)   (2.5 3-2B 3-11A 66.7) (2.5 3-2B 3-11B 66.7)
   (2.5 3-3A 3-7A 66.7)  (2.5 3-3A 3-7B 66.7)  (3.7 3-3A 3-9 100.0)
   (1.2 3-3A 3-11A 33.3) (1.2 3-3A 3-11B 33.3) (2.5 3-3B 3-7A 66.7)
   (2.5 3-3B 3-7B 66.7)  (3.7 3-3B 3-9 100.0)  (1.2 3-3B 3-11A 33.3)
   (1.2 3-3B 3-11B 33.3))))
```

[5] Here, as well as in the example of the following function, *n* is only three, resulting in a branch with a very simple level composition. Branches with a higher *n*, being more interesting in this respect, were too large to serve as examples.

Branch Contents 2

Description: Displays the level composition of a branch as well as the contents of the individual levels, returning SC names and weighted %REL_n values.

Usage: Branch-Contents-2 *SC1 SC2 n*

Details: The function gathers all %REL_n comparisons in branch *n* of the RECREL comparison between *SC1* and *SC2*. The level composition can be inferred from the indentation of the result. The sole level-*n* comparison is the leftmost one at the top. The comparisons at each of the lower levels *n-1*, *n-2*,...2 are placed stepwisely to the right. Several level-two comparisons may be placed on a single row.

Each sublist in the result contains two SCs and a weighted %REL_n value. The values are rounded to one decimal place.

Limits for the arguments: (1) 2 ≤ *#SC1,#SC2* ≤ 7, (2) 2 ≤ *n* ≤ 5, (3) *n* ≤ *#SC1,#SC2*.

Example: (Branch-Contents-2 5-1 5-35 3) ***returns***

```
Branch 3 in the RECREL comparison between SCs 5-1 and 5-35.
The entries in each sublist: two SCs and a weighted %RELn value.

(((5-1 5-35 90.0)
  ((3-1 3-7A 4.9)    (3-1 3-7B 4.9)    (3-1 3-9 7.4)     (3-1 3-11A 3.7)
   (3-1 3-11B 3.7)   (3-2A 3-7A 1.6)   (3-2A 3-7B 1.6)   (3-2A 3-9 4.9)
   (3-2A 3-11A 1.6)  (3-2A 3-11B 1.6)  (3-2B 3-7A 1.6)   (3-2B 3-7B 1.6)
   (3-2B 3-9 4.9)    (3-2B 3-11A 1.6)  (3-2B 3-11B 1.6)  (3-3A 3-7A 1.6)
   (3-3A 3-7B 1.6)   (3-3A 3-9 3.7)    (3-3A 3-11A 0.4)  (3-3A 3-11B 0.4)
   (3-3B 3-7A 1.6)   (3-3B 3-7B 1.6)   (3-3B 3-9 3.7)    (3-3B 3-11A 0.4)
   (3-3B 3-11B 0.4))))
```

Branch Value

Description: Returns the final value of a branch.

Usage: Branch-Value *SC1 SC2 n*

Details: The function determines the weighted values of all %REL$_n$ comparisons in branch *n* of the RECREL comparison between *SC1* and *SC2*, updates higher-level values with lower-level ones and derives a final branch value. The result is rounded to one decimal place.

Limits for the arguments: (1) 2 ≤ #*SC1*,#*SC2* ≤ 7, (2) 2 ≤ *n* ≤ 5, (3) *n* ≤ #*SC1*,#*SC2*.

Example: (Branch-Value 5-1 5-35 3) ***returns***

```
Branch 3 of the RECREL comparison between SCs 5-1 and 5-35.
Branch value: 57.0.
```

■ BIBLIOGRAPHY OF CITED WORKS

Abbreviations:

JASI: Journal of the Arnold Schoenberg Institute
JMT: Journal of Music Theory
ITO: In Theory Only
MA: Music Analysis
PNM: Perspectives of New Music

Alphonce, Bo
1974 *The Invariance Matrix*. Ph.D. dissertation, Yale University.

Beach, David W.
1979 'Pitch Structure and the Analytic Process in Atonal Music: An Interpretation of Theory of Sets.' *Music Theory Spectrum* 1, 7-22.

Brinkmann, Reinhold
1969 *Arnold Schönberg: Drei Klavierstücke Op. 11. Studien zur Frühen Atonalität bei Schönberg*. Wiesbaden: Franz Steiner Verlag GMBH.

Castrén, Marcus
1989 *Joukkoteorian peruskysymyksiä*. Helsinki: Sibelius Academy.

1990 '%REL_n and T%REL.' Unpublished manuscript.

Chapman, Alan
1981 'Some Intervallic Aspects of Pitch-Class Set Relations.' *JMT* 25/2, 275-90.

Cherlin, Michael
1991 'Dramaturgy and Mirror Imagery in Schönberg's *Moses und Aron*: Two Paradigmatic Interval Palindromes.' *PNM* 29/2, 50-71.

Christensen, Thomas
1987 'Schoenberg's Op. 11, No. 1. A Parody of Pitch Cells from Tristan.' *JASI* 10/1, 38-44.

Clough, John
1983 'Use of the Exclusion Relation to Profile Pitch-Class Sets.' *JMT* 27/1, 181-202.

Ericksson, Tore
1986 'The IC Max Point Structure, MM Vectors and Regions.' *JMT* 30, 95-111.

Forte, Allen
1964 'A Theory of Set-Complexes for Music.' *JMT* 8/2, 136-83.

1972 'Sets and Nonsets in Schoenberg's Atonal Music.' *PNM* 11, 43-64.

1973a *The Structure of Atonal Music.* New Haven and London: Yale University Press.

1973b 'Basic Interval Patterns.' *JMT* 17/2, 234-73.

1982 'The Magical Caleidoscope. Schoenberg's First Atonal Masterwork, Opus 11, Number 1.' *JASI* 5, 126-68.

1985 'Pitch-Class Set Analysis Today.' *MA* 4, 29-58.

1988 'Pitch-class Set Genera and the Origin of Modern Harmonic Species.' *JMT* 32, 187-334.

Hoover, Mark
1984 'Set Constellations.' *PNM* 23, 164-79.

Hämeenniemi, Eero
1983 'Joukkoluokkien intervallikot.' *Sibelius Academy Yearbook SIC*, 30-43.

Isaacson, Eric
1990 'Similarity of Interval-class content Between Pitch-class Sets: The IcVSIM Relation.' *JMT* 34, 1-28.

1992 *Similarity of Interval-class content Between Pitch-class Sets: The IcVSIM Relation and its Application*. Ph.D. dissertation, Indiana University.

Lewin, David
1959 'Intervallic Relations Between Two Collections of Notes.' *JMT* 3/2, 298-301.

1977 'Forte's Interval Vector, My Interval Function and Regener's Common-Note Function.' *JMT* 21/2, 194-237.

1979-80a 'Some New Constructs Involving Abstract Pcsets, and Probabilistic Applications.' *PNM* 18, 433-44.

1979-80b 'A Response to a Response: On Pcset Relatedness.' *PNM* 18, 498-502.

1987 *Generalized Musical Intervals and Transformations*. New Haven: Yale University Press.

Longley, Dennis, and Shain, Michael
1982 *Dictionary of Information Technology*. London and Basingstoke: The Macmillan Press Ltd.

Lord, Charles H.
1981 'Intervallic Similarity Relations in Atonal Set Analysis.' *JMT* 25, 91-111.

Morris, Robert
1979-80 'A Similarity Index for Pitch-Class Sets.' *PNM* 18, 445-60.

1982 'Set groups, Complementation, and Mappings among Pitch-class Sets.' *JMT* 26/1, 101-44.

1987 *Composition with Pitch-Classes*. New Haven: Yale University Press.

Neumeyer, David
1986 *The Music of Paul Hindemith*. New Haven and London: Yale University Press.

Perle, George
1977 *Twelve-Tone Tonality*. Berkeley: University of California Press.

1981 *Serial Composition and Atonality. An Introduction to the Music of Schoenberg, Berg and Webern.* University of California Press, Berkeley, Los Angeles, London. 5th ed.

Rahn, John
1979-80 'Relating Sets.' *PNM* 18, 483-97.

1980 *Basic Atonal Theory*. New York: Longman.

1989 'Toward a Theory of Chord Progression.' *ITO* 11/1-2, 1-10.

Regener, Eric
1974 'On Allen Forte's Theory of Chords.' *PNM* 13, 191-212.

Rogers, David W.
1992 A seminar handout describing three similarity measures. Eastman School of Music.

Schoenberg, Arnold
1975 *Style and Idea. Selected Writings of Arnold Schoenberg*. Edited by Leonard Stein. Trans. by Leo Black. London: Faber & Faber.

Solomon, Larry
1982 'The List of Chords, Their Properties and Use in Analysis.' *Interface* 11/2, 61-107.

Straus, Joseph N.
1990 *Introduction to Post-Tonal Theory*. New Jersey: Prentice Hall.

Teitelbaum, Richard
1965 'Intervallic Relations in Atonal Music.' *JMT* 9, 72-127.

Wittlich, Gary
1974 'Interval Set Structure in Schoenberg's Op. 11, No. 1.' *PNM* 13, 41-55.

ISBN 952-9658-32-X
ISSN 0788-3757
HAKAPAINO OY
HELSINKI 1994