DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA 91125

IMPLICIT PRESUPPOSITIONS: AN EXERCISE IN MULTIDIMENSIONAL SCALING AND HIERARCHICAL CLUSTERING

William L. Faust, Katherine Faust, and W. T. Jones

CALIFORNIA INSTITUTE OF TECHNOLOGY · THE TRUTH SHALL MAKE YOU FREE ·

HUMANITIES WORKING PAPER 83

May 1983

ABSTRACT

This working paper, like Humanities Working Papers 66 and 75, of which it is a further development, has two main aims. The first of these is to resolve a particular problem in art history. For this purpose the data already studied in Working Papers 66 and 75 are reanalyzed by means of multidimensional scaling and hierarchical clustering procedures, with results that support our earlier conclusion that sixteenth century Mannerism is best understood as an exaggeration of the High Renaissance style rather than as a distinct school. Our second aim, which in this paper takes precedence over the first, is to demonstrate to humanists that the quantitative methods of the social sciences can be used effectively to deal with some of the problems with which humanists are characteristically concerned, by replacing unresolved difference of opinion by judgments based on public procedures.

Though this is a joint paper, the text is chiefly the responsibility of a psychologist and an anthropologist; the explanatory comments and Discussion section are chiefly the responsibility of a philosopher. Thus the authorship of the paper reflects the kind of cooperation between social scientists and humanists that we are recommending.

IMPLICIT PRESUPPOSITIONS: AN EXERCISE IN MULTIDIMENSIONAL SCALING AND HIERARCHICAL CLUSTERING+

William L. Faust, Katherine Faust, and W. T. Jones

All humanists agree that The Prelude is a romantic poem and that The Rape of the Lock is not. But what are the features of the former that make humanists confident that it is romantic? And are those features the same as those that make us confident that Endymion and "To Autumn" are romantic poems? And what about Don Juan -- in what respects, if any, is it "romantic"? What about "Manfred"? Is it more, or less, romantic than Childe Harold? Such questions tend to lead to nonterminating disagreements because the notion of a school (alternatively, a style, a genre) is not well understood. How is romanticism bounded? Or, for that matter, how is phenomenology bounded, or structuralism, or behaviorism, or post-impressionism? We have devised a method by which we believe the notion of school can be clarified, thus making it possible to transform static confrontational disagreements about schools and school membership into ongoing problem-solving.

+We are much indebted to the following friends and colleagues for comments on Working Paper 75: James S. Ackerman, Brian Barry, John F. Benton, Judson Emerick, Margaret S. Faust, David Goodstein, George Gorse, Molly Mason Jones, Morgan Kousser, Oscar Mandel, Peter Manning, Jerome J. McGann, George W. Pigman, Aimèe Price, Alan Schwartz, Mary Martha Ward, Robert R. Wark, and Charles Young.

To demonstrate how this method can help terminate such disagreements, in Working Paper 75 (Jones, Faust, Faust and Jones, 1982) we examined a well-known art historical disagreement about Mannerism: Is Mannerism a "phase" of the High Renaissance? or is it a distinct school? Using data we had accumulated in earlier studies (presented in Humanities Working Paper 66 [Jones, Faust, Faust and Jones, 1981]), and analyzing these data by simple statistical procedures, we concluded that those art historians are right who regard Mannerism as a deviation from, or exaggeration of, some features of the High Renaissance style. But we also pointed out that one of the advantages of the method we have used is that, if this conclusion is disputed, more rigorous analysis of the data would be possible which would sharpen the definition of a school and identification of school membership, and we mentioned multidimensional scaling as one such possibility. In this paper we present the results of such a reanalysis of the original data. Here again, as in Working Paper 75, we are less interested in this particular art-historical disagreement -- though we believe our conclusions are not without interest -- than in demonstrating the wider applicability and relevance of the method used.

In emphasizing the importance of statistical analysis it is not our recommendation that humanists convert themselves into quantitative social scientists and become adept at multidimensional scaling. Rather, we propose a division of labor. If humanists are but willing to gather empirical data, they can tackle questions that are as relevant to the "boundary" and "school-membership" problems that interest them as the data used in Working Paper 75 are relevant to the art-historical problem discussed there. Once humanists assemble the appropriate data others can be found to perform a statistical analysis for them, and these analyses will yield knowledge about creative products not so much as dreamed of by those content to rely on a mere impressionistic "think so."

Inasmuch as the statistical analysis employed in this paper is formidable and yet we hope to demonstrate the usefulness to humanists of this method of tackling problems in the humanistic disciplines, we have adopted a somewhat unusual format for this paper. The text of the paper appears on successive right-hand pages. On left-hand pages, opposite points which may be difficult for humanists to follow, we provide explanatory comments in what we hope is nontechnical language.

Although the right-hand text does not assume familiarity with multidimensional scaling and hierarchical clustering, it is primarily intended for readers who feel comfortable using and comparing different formal models in analyzing data and who understand statistics. Even so we hope that humanists will not rely exclusively on the left-hand pages, but that they will also try the right-hand pages. As an inducement to them we have included in the text more explanation and elaboration than is strictly necessary for social scientific readers.

The special format begins here, with the humanist descant on the left-hand pages and the melody running successively on the right-hand pages.

INTRODUCTION

This Working Paper fits in the series of Working Papers (Humanities Working Papers 66 and 75 and Social Sciences Working Papers, Jones, Faust, Faust and Jones, 354, 355; Faust, Faust, Jones and Jones, 357) and articles by Jones (1970, 1972, 1973, 1976, 1980) on nonterminating disagreements and the differing "implicit presuppositions" which occasion such disagreements. However, the primary aim of this Working Paper, like that of Humanities Working Paper 75, of which it is a further development, is less to resolve a particular issue than to demonstrate to humanists that the quantitative methods of the social sciences can be used effectively to deal with some of the problems (those seemingly unresolved differences of opinion) with which humanists are characteristically concerned.

In Humanities Working Papers 66 and 75, we proposed (1) that the paintings of the earlier Italian period (1500-1515) can be grouped together as similar (that is in a "Renaissance School") since they exhibit considerable family resemblances, (2) that the Italian painting of the later period (1545-1560) show much less family resemblance among themselves and that many show striking family resemblance to the Renaissance family.

In this working paper we will reanalyze the same data which we analyzed in Working Paper 75. One goal of this reanalysis is to construct a model of the similarities among Italian painting during the 1500s which will represent the possible structure or pattern of similarities among the paintings in a way which makes this structure

**If humanists can overcome an initial, and natural, resistance to a vocabulary that includes such expressions as those used in this paragraph, there is nothing difficult about any of this. Suppose we want to know whether the male descendants of Queen Victoria through King Edward VII resemble each other physically more than they resemble her male descendants through Emperor William II and whether the latter group of descendants resemble each other more than they resemble the former group. This is equivalent to asking whether her English and German descendants form two (physical) families or only one, and it therefore corresponds to the question raised in the paper itself about Italian Renaissance paintings.*

***It is surely obvious that if we use physical dimensions (height, weight, etc.) to compare Victoria's descendants we may get a different pattern, or clustering, from the one we would get if we used intellectual or moral criteria. And of course even if physical criteria work well for Victoria's descendants we couldn't know, for sure, that physical criteria are <u>in general</u> good measures of family resemblance, until we made similar studies of the descendants of other monarchs (e.g., Maximilian of Hapsburg) and notables (e.g., John D. Rockefeller).*

more apparent than when we simply scan the data.* In any search for patterns, the properties of the paintings, the way these properties are assessed, and the analytic model used to analyze such data will all contribute to the structure which is developed.**

The reanalyses of the Italian paintings presented in this paper will use the same four dimensions which were used in Working Papers 66 and 75 on which to assess the similarity among the paintings. These four dimensions have been used in a number of other studies (see also Social Science Working Papers 354, 355, and 357).

The reanalyses of that data will use two different models of analysis: multidimensional scaling and hierarchical clustering. This exercise will support our earlier constructions and thereby provide some evidence of generality across methods of analyzing the data. In addition, the reanalysis will demonstrate the power of a more rigorous method in dealing with questions of the kind presented in Working Papers 66 and 75.

For the purposes of multidimensional scaling and hierarchical clustering we first assess the similarity-dissimilarity between each pair of paintings. Then the multidimensional scaling procedure develops a spatial representation of these similarities by plotting the paintings as points in Euclidean space much as the stars in the sky are projected on the ceiling. (See Figure 1, p. 79, for such a plot.) In the plot the similarity-dissimilarity among items (here the paintings) is represented in Euclidean space in such a way that items which are similar are close together on the plot, and items which are less

*We will not get very far in the study of Victoria's descendants if we rely on impressions of "look alike" and "look different." We want some objective measure of similarity and difference. Let us therefore settle on some physical features with respect to which we will compare the English and German descendants, and suppose, after some discussion, we agree to use four such features -- height, weight, head size, and hair color. There is no special problem about obtaining data on the first three measures -- a tape measure and a scales will do, and the "reliability" of the results of these measurements can be expected to be high: we can expect raters not to differ very much among themselves in the ways in which they read off height and head size from a tape measure and weight from a scales. But of course we shall test that expectation by comparing the agreement among raters. However, there is an initial problem about the measurement of hair color. Suppose we decide to have the descendants' hair color rated on a tonality scale from light to dark (eliminating the problem of rating differences in tint, like Titian red or auburn). Even so it is unlikely that all raters will locate a particular hair color at exactly the same mark on the light/dark scale. Hence we must test the reliability of hair ratings before we are able to use the results in our study of the English and German descendants of Victoria. (This corresponds to the preliminary testing of reliability done on the four dimensions, as reported in Working Paper 67.)

Let us suppose that ratings on the hair color scale prove as reliable as the four dimensions used in our earlier Working Papers. We similar are further apart. The dissimilarity is proportional to the distance between items on the plot. We are proposing that groups of paintings which have a "family" resemblance will be in close proximity in the spatial representation, whereas those which do not share a "family" resemblance will be dispersed throughout the space.

PROCEDURE

The data analyzed in this paper were obtained in three studies that have been reported in Humanities Working Papers 66 and 75. Those three studies analyzed eight Italian paintings from period 1500-1515 and ten from after 1525. Subjects (179 in the three studies) rated the paintings on scales developed to measure* different implicit presuppositions. Each painting was rated on four scales, each scale measuring a different presupposition. Each scale had eleven possible scale values.

Descriptions of the subjects, the procedures and the scales are given in Humanities Working Papers 66 and 75.

RESULTS

In this section we shall consider, first, whether the four scales are appropriate components for the measure of similarity-dissimilarity among paintings. Then, we shall discuss how the values on these four scales are combined to give a measure of similarity-dissimilarity. Next, the results of the multidimensional scaling will be presented and the statistical adequacy of the plots which will be are still not ready to begin our study. Though many of Victoria's English and German descendants have died or left no forwarding addresses, the subset of surviving descendants is so large that it would be an enormous undertaking to measure them all. We will therefore ask a group of genealogists, and perhaps the Garter King at Arms, to list for us a number of individuals whom they regard as "typical" or "representative" descendants, and we will draw eighteen descendants (eight English and ten German) from this list for study. (This corresponds to the procedure, described in Working Paper 67, by which eighteen sixteenth century Italian paintings were chosen for study.)

And now let us suppose that we have measured these eighteen descendants on our four scales and so have learned the height, weight, head size and hair color of each. These measurements are the data which we will use to determine whether the English and German descendants form two, or only one, families.

*The notion of redundance is important for the purposes of this paper and -- happily for us -- it is easily understood. Suppose we had chosen girth, instead of height, as one of the four measures on which to compare Victoria's English and German descendants. Weight and girth are likely to "go together" -- that is, to be positively correlated. Heavy people are likely to be big around the middle; light people, to be small around the middle. That being the case, if we knew the distribution of weights among Victoria's English and German developed will be considered. Finally, the results of the hierarchical clustering will be presented and the statistical adequacy of the clustering will be considered. The logical adequacy and the empirical relevance of the results of both procedures will be considered in the Discussion section of the paper (pp. 55, ff.).

Each of the eighteen paintings was rated on four scales and for each scale the median of these ratings was computed. The pattern of these medians was analyzed in Humanities Working Paper 75 and we repeat the tables, from that paper, which present the medians. Table 1 (see p. 72) presents the medians for each of the paintings from the early period for each scale, while Table 2 (see p. 72) presents the medians for each of the eight paintings from the later period for each scale.

Intercorrelations Among Scales

A first question is, Are the four scales independent measures? The wordings of the four scales are different, but do they measure different characteristics or do two or more of the scales measure the same characteristics? If two or more scales measure the same characteristic, then one of those scales can do the work for all of the scales measuring the same characteristic -- the others are redundant. If some of the scales are redundant use of more than one would give multiple weight to that dimension in the measures of similarity to be derived from those scales. The intercorrelation among the four scales will provide information which will help us make the decisions concerning redundancy.*

descendants, we could infer, with a reasonable probability, what the distribution of their girths would be. Since the second measure (girths) would tell us little more than the first measure (weight) tells us, it would be redundant. We would not need both. That is why, having chosen weight as a measure, we did not choose girth, but height. For though weight and height are pretty obviously correlated, the correlation will be modest. That is, there are many short people who are heavy and many short people who are light, and also tall people who are heavy and tall people who are light. Head size and hair color are not highly correlated with each other or with height or weight. Thus, since all four measures give us independent information about the physical similarities and dissimilarities of Victoria's descendants we can use all four in our study.

*Suppose that you and a friend decide to wager 50 cents on each toss of a coin. You take a coin from your own pocket. You suggest that he can call heads or tails and he tells you to flip the coin. You flip the coin; your friend calls tails while the coin is in the air and the coin lands tail side up. He calls tails on the next four successive tosses and the coin falls tails side up each time, so that he has won five straight times. You begin to wonder whether something has gone wrong. Possibly your friend has managed to substitute a coin of his own, one that is not evenly balanced, for your honest coin; possibly he can influence the fall of coins; possibly he has extrasensory perception. Another possibility of course is that the

The intercorrelations were computed[1] by the Pearson Product-Moment correlation, and these are presented in Table 3 (p. 73). There is a significant[*] correlation between scale 2 (needs little decoding/needs much decoding) and scale 4 (stability/change) and between scale 3 (whole/part) and scale 4.

coin is fair and that your friend has simply had a run of good luck. Which is it?

We can develop an abstract model to test whether the coin is fair. In such a model, we expect that on each toss of a fair coin heads and tails have an equal likelihood. That is, on each toss heads would have a probability of 1/2 (.5) and tails would have a probability of 1/2 (.5). The probability of getting tails on two successive tosses would be the probability of tails on the first toss (1/2) times the probability of tails on the second toss (1/2); 1/2 times 1/2 = 1/4 or .5 times .5 = .25. The probability of a <u>fair</u> coin coming up tails on five successive tosses is 1/2 times 1/2 times 1/2 times 1/2 times 1/2 = 1/32 = .031 or about 3 times in 100 by chance. Another way of putting this is to say that in 100 runs of coin tossing of the kind described one could expect a sequence of five straight tails to occur three times by chance.

Is that enough to make you call off the game? The computation we have just made does not, and cannot, tell you what to do. It tells you only how likely it is that the coin is fair, thus giving information that will help you make a more rational decision than you could make if you had nothing to go on but your impression of your friend's honesty. Remember that the hypothesis tested by the statistical analysis is that the coin is fair, and the computation has shown that, in the long run, if the coin is fair the observed sequence of five tails would occur three or less times in 100 tosses (p = .031). What degree of unlikeliness should make you quit the game? The

probability value which you select is called the criterion of significance. In the behavioral sciences, the most frequently used criterion of significance is fewer than 5 times in 100, written probability less than .05 ($p < .05$)." This is an arbitrary choice and should be adjusted whenever there are good reasons to select some other value. Since in this study we have no good reason to choose any other criterion of significance we shall adopt $p < .05$ and get on with substantive decisions.

We asked how likely it was that the observed correlations between the various scales would have occurred if the scales are independent of each other (which is equivalent to asking how likely it is, if the coin is fair, that the observed sequence of tails would occur). The computation showed that in the case of two of the correlations -- that of scale 2 with scale 4 and that of scale 3 with scale 4 -- the observed correlations would have occurred fewer than 5 times in 100 ($p < .05$). This compares with our computation (in the example) that the run of five tails would occur 3 times in 100 ($p = .03$). Hence, using the .05 rule cited above, we rejected the hypothesis that these two scales are independent. Since all the other correlations -- scale 1 with scale 2, scale 1 with scale 3, scale 1 with scale 4, scale 2 with scale 3 -- were more likely to occur ($p > .05$), we concluded that these scales are independent of each other.

In our illustrative study of Victoria's descendants it was unnecessary to make the computations reported in the text: we have all

The square of the correlation (r^2) can be used as measure of the proportion of the variance among the scores of the dependent variable which is associated with variance among the scores of the independent variable. In this instance, significant negative correlation between scale 4 and scale 2 ($r = .61$, $r^2 = .37$) can be interpreted in terms of the percentage of the variances in scale 4 which is associated with variance in scale 2, e.g. 37 percent of the variance among scale 4 values is associated with variability in scale 2 values. The significant correlation between scales 3 and 4 is $r = .52$ and $r^2 = .27$. That is, 27 percent of the variance in scale 4 values is associated with variability in scale 3 values. This is a tolerable amount of redundance between two scales.*

The multiple correlation using scales 2 and 3 against scale 4 indicates that even the two scales, 2 and 3, together do not replace scale 4.

had so much experience with physical characteristics like weight, height and girth that we have a good idea which are strongly and which are only weakly correlated. But the four measures used in the real study of paintings are "new": nobody has had enough experience with them to know, without making the computation, whether they are strongly or weakly correlated, or whether they are independent of each other.

Humanist readers can safely take the "Pearson-Product-Moment" paragraph on faith -- the computation described is the sort of thing which, in any cooperative study with social scientists, humanists can expect to have done for them. The important point, for humanists and social scientists alike is the result of the computation, viz. that the four dimensions used in the study are not strongly correlated. This is important because the data from all four measurements are to be combined into a single measure. Obviously had any two, e.g. weight and girth, been strongly correlated, combining these two into a single measure with head size and hair color would have skewed the results -- it would have been equivalent to counting weight twice, with the result that descendants whose weights were close together would seem to resemble each other more than descendants whose hair color was similar.

*The median in any array is the mid-point in that array. Suppose, for instance, that we asked fifteen raters to rate each descendant's hair color on an 11-interval scale from "very light" to "very dark," with the following results for descendant a:

```
                   x  xx       x  x
                x  x  xx   x   x  x  x  x        x
        very   |--|--|--|--|--|--|--|--|--|--|--|  very
        dark    1  2  3  4  5  6  7  8  9 10 11   light
```

The median rating of descendant a's hair color is in interval 4, because seven raters have rated his hair color as darker, and seven have rated his hair color as lighter, than this rating.

**It is less important for humanists to understand "iterative normalization" -- the procedure described in the text -- than to understand why this procedure was used, i.e., what it accomplished for us. We will therefore concentrate here on explaining its purpose.

So far, in the illustrative study of Victoria's descendants, we can compare each descendant's weight with the weight of every other descendant, thus learning how similar the English descendants are in respect to weight, as compared with the German descendants. And we can

Standardization of the Ratings on Each Scale

Multidimensional scaling and hierarchical clustering both require as input a measure of similarity-dissimilarity between each pair of items in the set being examined -- in this case, the eighteen sixteenth century Italian paintings. To create such a pairwise measure, we took the medians* for each picture on each of four dimensions (across the three studies). In previous papers we reported the medians as alphabetical letters corresponding to positions on the rating scales. Here we have transformed them into numbers from 1 to 11. (See Tables 1 and 2, p. 72.)

The median values on these scales must be in a comparable metric before we can enter those values into the multidimensional program for computation. Though the median values on the scales did not differ appreciably in distribution, the values of each scale were standardized.

In order to standardize the median scale values across the four dimensions, the column totals were set equal, and the individual median scale values were adjusted appropriately so as to sum to the total. This procedure, called iterative normalization** (Romney, Keiffer and Klein, 1973), has the effect of translating the median scale values into proportions of the dimension total. These standardized measures were used to calculate pairwise similarity-dissimilarity.

compare them with respect to their similarities and differences on each of the other scales. But for the purposes of multidimensional scaling we must combine the differences between each pair on all four measures into a single, composite number which captures the multidimensional qualities of the differences between pairs. In order to achieve that goal we must make the values on each scale comparable with the values on the other scales. Obviously, measurements in feet and in pounds are not in equivalent scales (as are, say, measurements in feet and in meters. Moreover, the range of height would probably not exceed two feet, whereas the descendants might differ by as much as 100 pounds. Further, our pairs might differ by four steps on the hair color scale and by many more pounds on the weight scale. Normalizing the scores on each of these scales permits us to sum the difference for each pair and then compare each pair's combined measure with the corresponding combined measure of every other pair.

**"Euclidean distance" may sound somewhat alarming. It merely means the summed distances between any two descendants on all four dimensions -- height, weight, eye-color, hair-color. If we had nothing to go on but impressionistic "look alikes" and "look differences" we could only say unhelpful things like "descendant a is a good deal heavier and quite a lot taller than descendant b, and he has rather darker eyes and much lighter hair." Now that we have standardized the medians on all four dimensions, we can do considerably better. Suppose we have the height, weight, head size and hair color results for three of the descendants, a, b, c. (The numbers in the matrix are the*

Similarity-Dissimilarity Measure: Euclidean Distance*

The method of deriving the similarity-dissimilarity values[2] between pairs of paintings is of special interest because there are various ways in which similarity can be measured, and the "goodness" of the outcomes in multidimensional scaling and in hierarchical clustering vary with the particular method used.

The similarity between the paintings on these four scales was calculated by computing the Euclidean distance between each pair of pictures. For each pair this Euclidean distance equals:

$$\text{Euclidean distance} = \sqrt{\Sigma \text{across 4 scales} \left(\begin{array}{c}\text{score on a scale} \\ \text{for one painting}\end{array} - \begin{array}{c}\text{score on same} \\ \text{scale for} \\ \text{other painting}\end{array} \right)^2}$$

standardized medians for these three among the many descendants.)

Descendant	height	weight	head size	hair color
a	5	6	6	7
b	4	4	8	6
c	7	8	6	3
	24	24	24	24

Note that this iterative normalization procedure makes the sum of all columns equal. Scores for descendant "d" have been placed in this table, though we shall not consider "d" until later. At this point, and using only the values for a, b and c, we can now compare pairs -- ab, ac, bc -- by adding together the differences between ab, ac, and bc on all four dimensions and taking the square root. Here we utilize the formula for Euclidean distance given on p. 22:

$$\underline{ab} = \sqrt{(5 - 4)^2 + (6 - 4)^2 + (6 - 8)^2 + (7 - 6)^2} = 3.16$$

$$\underline{ab} = \sqrt{(4 - 7)^2 + (4 - 8)^2 + (8 - 6)^2 + (6 - 3)^2} = 6.16$$

$$\underline{ac} = \sqrt{(5 - 7)^2 + (6 - 8)^2 + (6 - 6)^2 + (7 - 3)^2} = 4.89$$

Thus ab are more similar than ac, and ac are more similar than bc. Three descendants are easy to deal with, but remember that the subset of descendants that we are studying contains eighteen individuals, and each of these eighteen descendants must be compared with every other descendant on the (composite) measure of similarity-dissimilarity that

Nonmetric Multidimensional Scaling

The procedure of nonmetric multidimensional scaling gives a geometric, i.e., spatial, representation of the degree of similarity-dissimilarity between every pair of items in terms of distance between points. The greater the distance between two points in the spatial model, the less similarity between the two items represented by those points. Since the multidimensional scaling procedure used here reflects the rank order of similarity between pairs, pairs which are *we are using. This results in a large number of pairs, and the addition of each new pair alters the relations of all the previously plotted pairs. The task of making all of these adjustments would be most tedious if it had to be carried out by hand. A computer will perform the calculations quickly and without errors creeping in.*

**Here is another possibly forbidding phrase. But actually "not proportional to the metric interval" is easy to understand. (If any humanist readers are interested in why the distances between pairs are not proportional to the metric interval, we give an account at the end of this explication). Meanwhile, to understand the phrase, concentrate, by way of contrast, on what a plot looks like when the points on it are proportional to the metric interval. Any map is an example. If it is five times as far (say) from Los Angeles to San Francisco as from Los Angeles to Santa Barbara, then the interval on the road map between the point representing Los Angeles and the point representing San Francisco will be five times longer than the interval between the point representing Los Angeles and the point representing Santa Barbara.*

The computer is going to draw a picture (or plot, or map) that represents the similarities-dissimilarities among eighteen sixteenth century Italian paintings. In Figure 1, p. 79, each point represents a painting; in the illustrative example we are using on these left-hand pages, each point may be thought of as representing a descendant. The distance between any pair of points (paintings, descendants) represents how similar-dissimilar that pair of paintings (alternatively, descendants) are.

more similar will be closer together in the spatial model. However, the distances between points in the spatial model are not proportional to the metric intervals* -- that is why this is called "nonmetric" scaling.

Now consider two pairs of points, one of which is five time as far apart as the other pair is. We cannot conclude (as we could, if this plot were a map of California and two pairs of points were Los Angeles/San Francisco and Los Angeles/Santa Barbara) that the second pair of points is five times more similar than the first pair. But we can conclude that in the rank order of similar-dissimilar pairs the second pair is less similar than the former pair.

This may not sound wildely informative when we are dealing with no more than two or three pairs, but as pairs are added the results becoming increasingly relevant to the question we want to answer: Do Victoria's English descendants and German descendants form two distinct groups? For consider: if all of the English/English pairs resemble each other and all of the German/German pairs resemble each other but the English/German pairs are not similar, then the English and German descendants form two distinct families. However, if some English/German pairs are very similar, then it will look as if the English and German descendants are best regarded as a single family or perhaps an undifferentiated collection.

Humanists who are interested in learning why the points are not "proportional to the metric intervals" should read on; those who are not may stop here. Consider the difficulties we would encounter if we tried to develop a map (or plot, or spatial model) of the descendants' similarities-dissimilarities in which the intervals are metric. Start with a few points and a ruler-like spatial model, with equal intervals. In the example we have already used (see p. 23 above) suppose a is 3.16

units away from b and b is 6.16 units away from c. Thus:

```
  a         b                  c
  |--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
```

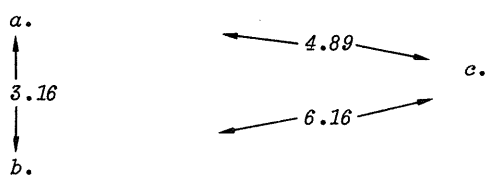
The distance between a and c along this ruler-line is 9.32 units but the similarity-dissimilarity measure gives the distance between a and c as 4.89 units. If we make ac equal to 4.89 units, we cannot keep ab equal to 3.16 and bc equal to 6.16 in a two dimensional straight-line representation of the relative distances.

We could, however, try another linear representation.

```
                    b        a          c
```

Here ab is 3.16 units and ac is 4.89 but bc is 8.05 units on the ruler scale although it should be 6.16 similarity-dissimilarity units.

Additionally we could present the similarity-dissimilarity distances in three dimensions in a triangular pattern



If there are many pairs, rather than the three of our example, we might not be able to present all the similarity-dissimilarity distances even in a three dimensional picture. Distances in Figure 1 (p. 79) do not try to reflect the metric units of the similarity-dissimilarity measure, but rather reflect the rank order of pairs from the most similar pair (closest together in the space) to the least similar (farther apart in the space): Thus, in our universe of three pairs -- ab, bc, ac -- we can say that ab is the most similar pair and

that ***bc*** *is the least similar (most dissimilar) pair:*

	Euclidean Distance	Rank order of similarity
ab	3.16	1
bc	6.16	3
ac	4.89	2

Only the triangular representation retains the metric distance. However, either the second linear or the triangular spatial representation of the relations among the three descendants retains rank order since a and b are represented as closer together than a and c, and a and c are represented as closer together than b and c. But the distances in the linear representation are certainly nonmetric.

The scaling procedure provides a model in which the rank order of the distances among the points approximate the rank order of the dissimilarities in the input data. If we were willing to have a model with a large number of dimensions we could perfectly represent the original data. However, since we are unlikely to be able to perceive the structure of a model in more than three dimensions, we choose a representation of fewer (usually two or three) dimensions which, though imperfect, allows us to easily see the structure in our data. Here we have chosen a two-dimensional model.

These similarity-dissimilarity measures were analyzed using the KYST multidimensional scaling program. (Kruskall, Young and Seery, 1973). Figure 1 (p. 79) presents the two-dimensional plot which that program developed.

Analysis of Multidimensional Scaling Outcome

Two questions need to be asked about the multidimensional scaling plot given in Figure 1.

(1) Are the distances between points in the plot a good fit to the similarities between items? To answer this question we ignore the distinction between Renaissance and later paintings. We simply ask, "Are the distance among the eighteen paintings in the multidimensional model a good fit to the similarities-dissimilarities among those paintings in the original data?" In order to answer this question, we shall report values of stress and gamma.

(2) Do the Renaissance and later paintings cluster together in ways that correspond to art historians' views? That is, do the

similarities-dissimilarities we have uncovered by the multidimensional scaling procedure make art-historical sense? In this second analysis we shall ask whether the eight Renaissance paintings are clustered together and whether the ten later paintings are clustered together. In order to answer this question we shall report the results of the Quadratic Assignment procedure.

Goodness of Fit

The goal of the nonmetric multidimensional scaling procedure is to present a configuration of points in a spatial model in which the rank order of distances between points are the best approximation to the rank order of similarity-dissimilarity between items. Frequently, only approximate solutions can be obtained.

"Goodness of fit" describes the degree to which the relationship between rank order of similarity-dissimilarity is maintained in the rank order of distances in the spatial array.

There are various ways of measuring the fit between the rank order of distances in the spatial configuration and the rank order of similarity-dissimilarity values. We will consider two of these, stress and gamma, in order to ascertain the fit between the spatial array in Figure 1 and the similarity-dissimilarity data obtained for the eighteen paintings.

In the example given above (p. 21), where we were dealing with only three similarity-dissimilarity pairs, ab, bc, and ac, the relative rank order of similarity-dissimilarity was equivalent to the rank order of distances in the second linear array and in a triangular spatial array. That is, "goodness of fit" was achieved for rank orders of similarity-dissimilarity values. Below, by way of contrast, we give an example of difficulties of fit. Let us suppose that we add a fourth descendant, d, to the three we have already studied. The addition of d gives us three more similarity-dissimilarity pairs, inasmuch as d must be compared with a, with b and with c. We repeat the earlier table for your convenience.

Descendant	height	weight	head size	hair color
a	5	6	6	7
b	4	4	8	6
c	7	8	6	3
d	8	6	4	8
	24	24	24	24

We now compute similarity-dissimilarity values for ad, bd and cd.

$$\underline{ad} = \sqrt{(5-8)^2 + (6-6)^2 + (6-4)^2 + (7-8)^2} = 3.74$$

$$\underline{bd} = \sqrt{(4-8)^2 + (4-6)^2 + (8-4)^2 + (6-8)^2} = 6.32$$

$$\underline{cd} = \sqrt{(7-8)^2 + (8-6)^2 + (6-4)^2 + (3-8)^2} = 5.83$$

Badness of Fit-Stress*

A badness of fit measure, called stress, can be computed -- the stress figures range from 0.0 to +1.0, and the larger the stress value (e.g., closer to 1.0) the worse the fit between the interpoint distances and the original similarity-dissimilarity values. The stress value for the plot in Figure 1 is .0894 -- a value which indicates very little stress. Alternatively phrased, the low stress value leads to the inference that the fit is quite good.

Here are all of the Euclidean distances and their rank orders:

	Euclidean Distance	*Rank order of similarity-dissimilarity*
ab	*3.16*	*1*
bc	*6.16*	*5*
ac	*4.89*	*3*
ad	*3.74*	*2*
cd	*5.83*	*4*
bd	*6.32*	*6*

And now let us attempt to locate the new results, along with the old ones, on a linear array. We shall use the rank order values in deciding a distance just as the multidimensional program does. Note that a is closest to b, and b is closer to c than to d. This could give the following linear array:

a   b   c   d

But a should be closer to d than to c, because ab is second and ac is third in the rank order of similarity-dissimilarity. On the other hand, b should be closer to c than to d because bd is fifth and bc is sixth in the rank order of similarity-dissimilarity.

Just how bad the fit is between the rank order of similarity-dissimilarity and the rank order of spatial distances can be seen in the following table:

	Rank order of similarity-dissimilarity	Rank order of distances between points in linear array
ab	1	1
bc	5	2
ac	3	3
ad	2	6
cd	4	4
bd	6	5

Since the two rank orders are not congruent, the straight-line representation given above is not a "good fit" to the similarity-dissimilarity data. A straight-line spatial array (map, model) is not a proper "picture" of the family resemblances among the four descendants, a, b, c and d. If we want a faithful picture some better spatial representation would be sought by a multidimensional program and a better fit can be found. But no exact fit can be found in a linear representation.

**Humanists will do well to accept "badness of fit" and "goodness of fit" on faith -- in the cooperative kind of inquiry we are proposing these tests are the responsibility of the social scientists. But it is important that humanists understand the results of making the tests. They show that the fit between the spatial array presented in Figure 1 (p. 79) and the similarities-dissimilarities of the eighteen paintings, as measured on the four dimensions used in the test, is very*

Goodness of Fit-Gamma*

Another way to look at the goodness of fit between the spatial configuration and the similarity-dissimilarity values in the data is provided by the Goodman and Kruskall gamma (Freeman, 1965). Gamma can also vary from 0.00 to +1.00, but in this case higher values indicate a better fit. The gamma for the multidimensional scaling spatial configuration presented in Figure 1 is gamma = +.845 again, leading us to infer that the fit is very good indeed.

So much for our answer to the first of the two questions posed on p. 32. We turn now to the second.

good. In other words, the picture reflects the similarity-dissimilarity measure quite nicely.

**Returning to our example of Queen Victoria's descendants: we want to know whether the English and German descendants form two distinct families. That is, do the English descendants cluster together more than similar sized groups composed of English and/or German descendants that we might arbitrarily select? And do the German descendants also cluster together in this way? We define "cluster together" in terms of distances in the spatial array -- the shorter the distances among items in the array, the greater the clustering of those items. We can define "cluster together" in this way because we now know, as a consequence of our tests by stress and gamma, that distances in the spatial array are congruent with the family resemblances we want to study.*

***One way to think of this procedure is to imagine a randomly selected subset of eight descendants from the set of eighteen descendants whose similarities and dissimilarities we are studying and to compute the average distance among this subset of eight. Were this done for all possible sets of eight descendants among the eighteen, some sets would be compact, and the average distance among them would be small. Other sets of eight randomly selected descendants would be scattered, and the average distance would be large. The question is, where in this distribution do our set of eight English descendants and our set of ten German descendants fall?*

Significance of Spatial Clusters

Figure 1 (p. 79) shows the positions of the earlier and the later paintings. The Renaissance paintings tend to cluster in the left half of the space; most are in the lower left quadrant. The later paintings are more evenly distributed across the other three quadrants. But are these spatial clusters closer together than would be expected by chance?*

Hypotheses about the relative degree of clustering or dispersion of a subset of items in relation to the whole group can be tested using the Quadratic Assignment Program (Hubert and Schultz, 1976).** This program provides a measure which can be translated into the average distance among all pairs of paintings in our set of eight, and another measure which can be translated into the average distance among all pairs of eight paintings among all possible sets of eight paintings in the eighteen. This procedure allows us to see where the Renaissance set of eight and the later set of ten fall in the distribution of all sets of eight paintings.

*Here again humanists need not trouble themselves about technicalities. In the division of labor we are recommending, "test statistics" are a responsibility of the social scientists. What concerns humanists are the results of the computation, which, in terms of our example, show that the English descendants form a distinct family and that the German descendants do not. That is to say, the distances that separate the points in the spatial array representing the English descendants are much less than the distances separating any arbitrarily selected subset of descendants, whereas the distances separating the German descendants are not.

The test statistic from the quadratic assignment procedure is Z.* For the average distance between Renaissance paintings Z = -2.03. Since these Renaissance painting represent a predicted, not an ad hoc, group, the quadratic assignment procedure leads to a strong inference (in a probability sense) that the clustering together of the eight Renaissance paintings is closer than would be expected by chance. The comparable Z value for the clustering of the later paintings is Z = .68, which leads to the inference that the later paintings are not clustered to a greater extent than would be expected from a chance selection of paintings. (See Table 4, p. 74.)

<u>Interpretation of the Axes of the Multidimensional Scaling Spatial Arrangement</u>

The configuration given in Figure 1 (p. 79) can be analyzed to discover how the paintings are arranged in the plot in relation to the dimensions which comprised the data. The PROFIT program developed by Chang and Carrol (1968) was used to relate the four implicit presupposition dimensions to the axes of the spatial arrangement shown in Figure 1 (p. 79). Figure 2 (p. 80) presents the vectors for the four dimensions. Dimension 4 has a vector very close to the horizontal. Painting to the left in Figures 1 and 2 are rated as more stable (at rest) while those to the right are rated as more in flux or change. Dimension 1 is slightly oblique but close to the vertical. Paintings in the upper part of the array are more outer and those in the lower part are more inner.

**So far we have studied the similarities-dissimilarities among pairs of paintings (descendants) by locating points in a spatial plot (a "picture") in which the distance between any two points represents the position of this pair of paintings in the rank order of similarity-dissimilarity of all pairs of paintings. This procedure has tended to confirm the conclusion we reached in Working Paper 75.*

Now we use a different procedure to analyze the same data. Here the data -- the similarity-dissimilarity pairs -- are to be located in a different kind of picture, one which will bring into focus other features of the pairs. Whereas the picture that results from multidimensional scaling looks very much like an ordinary map (though it isn't an ordinary map because the distances do not correspond to the distances on the area mapped), the picture that results from hierarchical clustering will look like an ordinary family tree (though it won't be a family tree because the groupings do not represent generational relationships -- fathers, sons, grandsons, uncles, first and second cousins. But, once again, the physical similarities-dissimilarities of pairs of descendants).

***To explain how hierarchical clustering is done and to illustrate what it accomplishes, let us consider the sums of the standardized medians for each pair of five descendants a, b, c, d, and e. In hierarchical clustering we are concerned -- as we were in multidimensional scaling (see for instance the table on p. 37) -- with the rank order of the similarity-dissimilarity pairs. But, instead of*

Dimensions 2 and 3 form obliques. Accordingly, paintings located in the upper left part of the array are rated as relatively surface and those in the lower right as relatively depth, while paintings in the lower left are rated as emphasizing whole and those in the upper right as emphasizing parts.

Of course, the location of the dimensions on the spatial plot in relation of the dimensions to the horizontal and vertical axes is a function of the program; it does not represent a real spatial orientation in the data, i.e., the whole page could be turned 90 degrees and the relations would hold.

Hierarchical Clustering

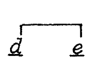
Geometric representation is not always the best model in which to represent the structure of similarities among items. Another way to look at the similarity-dissimilarity data between pairs is by means of a procedure known as hierarchical clustering.*

The hierarchical clustering model used here provides a tree diagram (dendrogram) in which classes are embedded within other classes so as to yield a hierarchical structure ranging from minimal clustering (in effect, zero clustering) in which each item is separate, to maximal clustering in which all items are associated in one cluster.**

Such solutions require that, as one proceeds from minimal clustering to maximal clustering, if any two items appear together in the same cluster, they cannot subsequently be separated and placed in different clusters. Such a hierarchical class structure loses some representing the rank order as distances in a spatial model, in the hierarchical clustering procedure we represent the rank order by different levels in a tree-like form, in which a very similar pairs are represented as being on the same level (as with two siblings in a genealogical tree) and less similar pairs occupy increasingly more remote levels (as with a father and a son, who are separated by one level in a genealogical tree and a grandfather and a grandson, who are separated by two levels).

Pairs	Euclidean Distances
ab	5
ac	15
ad	35
ae	37
bc	10
bd	30
be	32
cd	20
cd	22
de	2

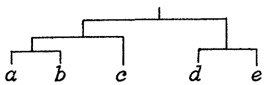
Inspection of the table above shows that de is the most similar pair. Accordingly, d and e are clustered (combined) first:



The next most similar pair is ab. Therefore a and b are next clustered:

This leaves descendant c to be clustered. Is c more similar to a and b or to c and d? Clearly, it is more similar to a and b than to c and d, because ca = 15 and cb = 10 (total 25), whereas cd = 20 and ce = 22 (total 42). Therefore c clusters with a and b:

Finally, since in this example we are dealing with but five descendants and all have been clustered, we complete the diagram by clustering all together.



*In our descendants example, the first question amounts simply to this: Is the hierarchical cluster in Figure 3 a reliable "picture" of the similarities-dissimilarities among the eighteen descendants, as these are measured by our (combined) four physical criteria: height, weight, head size, and hair color? The answer, it will be seen, is, Yes, the hierarchical cluster in Figure 3 (p. 81) is a reliable picture, but it is not as reliable as the "picture" developed by means of the multidimensional scaling procedure (Figure 2).

**Assuming, then, that Figure 3 (p. 81) is reliable, the second question is, What can we infer from it about the family resemblances among Victoria's English and German descendants? In particular, do the information pertaining to relative distance between items and between classes, but strongly hierarchical structures provide a great deal of information about the nature of the similarity between items.

The hierarchical clustering of the similarity-dissimilarity among the paintings we have studied was performed by the U-Statistic Hierarchical Clustering Program (ALPAIR) developed by D'Andrade (1978). Figure 3 (p. 81) presents the tree diagram developed from the outcomes of the clustering procedure.

Analysis of Results of Hierarchical Clustering

We shall raise, and answer, the same two questions posed early (p. 32) about the multidimensional scaling plot:

1. Is the hierarchical clustering of the eighteen items a good fit to the similarity-dissimilarity values among all eighteen paintings, disregarding any distinctions between Renaissance and later paintings? For this analysis we shall rely on gamma.*
2. Does the hierarchical clustering diagram cluster two groups: one group of eight Renaissance paintings and another group of ten later paintings? We shall use the quadratic assignment test to evaluate this question.**

1. Goodness of Fit -- Gamma

The goodness of fit between the cluster levels and similarity-dissimilarity data as measured by gamma is 0.658. Since values of gamma vary from 0 to 1.0, where 0 is a poor fit and 1.0 is a good fit, we infer from this gamma that the hierarchical clustering represents

English descendants resemble each other more than they resemble the German descendants? Do the German descendants resemble each other more than they resemble the English descendants? The answer (as far as the hierarchical cluster goes) is that the descendants do indeed cluster into two groups, but that there are English and German descendants in both groups. This is the case because, amongst the strongest resemblances, there are four English/German pairs.

*This is the case because the first pairs to be clustered are always those that are most similar (thus, in the example on p. 47 we clustered de first and then ab; c came in later, at the second level). Since clustering at each subsequent level requires less and less similarity, the upper levels of a tree such as that in Figure 3, p. 81, represents relatively little similarity. Thus the more inclusive groups that include items in these upper levels do not represent much family resemblance among the items so grouped.

the similarity-dissimilarity among the paintings quite well, although not as well as the multidimensional scaling represented the same similarity-dissimilarity data.

2. Description of Hierarchical Clusters in Terms of Renaissance Paintings and Paintings From the Later Period

If the hierarchical clustering procedure were to yield two schools of paintings, a Renaissance school and a school of the later paintings, as some art historians maintain, then at the level in Figure 3 (p. 81), where there are only two clusters, all or most of the Renaissance paintings would comprise one cluster and all or most of the paintings from the later period would comprise the other cluster. Actually, at that level six Renaissance paintings and three paintings from the later period are grouped in one cluster, and two Renaissance paintings and seven paintings from the later period are grouped in the other cluster. However, though we can thus say that one group is predominantly composed of paintings from the later period, this level of clustering, where there are only two clusters, does not require a great deal of similarity among members of each cluster.*

More significantly, therefore, is the lower level at which the paintings are grouped in pairs. Here there are four instances where a Renaissance painting is paired with a painting from the later period (1, 14; 4, 15; 7, 17; 6, 16) two instances where both paintings are from the later period (11, 13; 10, 12), and one instance where both are Renaissance paintings (5, 8). It is quite clear that at the level of clustering in pairs, where the requirements for similarity between

**Do not be exercised by "Z = -1.6." The hypothesis that is tested statistically is usually the contradictory of the hypothesis that is of interest to us. Thus if the hypothesis we wish to support is that English and German descendants form two distinct physical groups, then the hypothesis that we test statistically is that the English and German descendants do not form two distinct physical groups (see p. 11 above for an explanation of the logic of this seemingly inverted approach). If we cannot reject the hypothesis that almost any random set of descendants selected from the eighteen descendants, cluster as tightly as the English descendants cluster and the German descendants cluster, then we do not have good evidence (at least from these data) that the procedures of hierarchical clustering identified two distinct family groups.*

This is the last comment on the technical exposition. With the Discussion section that begins on p. 55; we resume the regular format.

items is greater than it is in the clusters comprising more items, the clustering procedure did not pair Renaissance paintings with other Renaissance paintings nor paintings from the later period with other paintings from that period.

Significance of the Grouping of the Renaissance and of the Paintings From the Later Period

Earlier in this paper (pp. 38-40), the quadratic assignment procedure was used to evaluate the clustering of the Renaissance paintings and the clustering of the paintings from the later period as they were represented in the plot derived from multidimensional scaling. This procedure can also be used to evaluate the groupings of these paintings in the hierarchical clusters shown in Figure 3.

The clustering of the Renaissance paintings gives a $z = -1.6$;* and the clustering of the paintings from the later period give a $z = .32$.

The probability associated with these two values leads us to infer that neither group of painting forms a coherent group. The eight paintings from the Renaissance school are not clustered more than might be expected by choosing, at random, any subset of eight paintings from our set of eighteen paintings; the ten paintings from the later period are not clustered more than might be expected by choosing, at random, any subset of ten paintings from our set of eighteen paintings.

This concludes the technical exposition and with it the need for the special right-hand/left-hand format that we have been using. We therefore revert to the regular format.

DISCUSSION

In this final section of the paper, though we focus attention on the light that multidimensional scaling and hierarchical clustering throw on the Renaissance/Mannerism debate, our larger aim is to demonstrate the relevance of social-scientific procedures to a wide range of humanistic inquiries.

Integration of the Results of Multidimensional Scaling and Hierarchical Clustering

First, consider the results of the two procedures separately:

The hierarchical clustering procedure developed two clusters -- one cluster predominantly composed of Renaissance paintings and the other cluster predominantly composed of paintings from the later period. But, since the pairing of paintings did not represent a pairing of Renaissance paintings with Renaissance paintings and paintings from the later period with paintings from the later period, and since the clustering procedure demands that once a pair is joined it must remain together in more inclusive clusters, the hierarchical clustering procedure did not significantly group together Renaissance paintings with Renaissance paintings nor paintings from the later period with paintings from the later period.

The most striking feature of the Renaissance cluster is its distribution along D-4 (rest/change). All the Renaissance paintings save one were rated as characterized by rest, whereas only one of the later paintings was rated in this way. (See Figure 2 (p. 80) and the accompanying text). Distribution along D-1 (inner/outer) is almost as strongly dichotomized: only two of the Renaissance paintings were rated as outer, and only two of the later paintings were rated as inner.[3] If we had had the eighteen paintings in our study rated on these two dimensions alone, we would certainly have concluded that those art historians are correct who maintain that High Renaissance and Mannerism are two distinct schools. It is possible, we think, that art historians may selectively attend to different features of paintings. Those for whom the features captured by D-4 and D-1 are prominent are likely to conclude that Mannerism is a distinct school, whereas those who selectively attend to the features captured by D-2 and D-3 are likely to conclude that Mannerism is not a distinct school but only an exaggeration of the Renaissance style. In a word, the art-historical disagreement may result from differential weighings of certain features of the paintings, some historians taking these, and others those, as the leading features. We suspect that many seemingly intractable disagreements -- and by no means only in art history -- can be dissolved in this way.

In contrast, the multidimensional scaling procedure, which adjusts distances between every pair of points in the plot in a way that takes account, so far as possible, of the distances among all of the pairs, found a single Renaissance cluster and at the same time provided a better fit to the similarity-dissimilarity values.

The multidimensional scaling procedure and the hierarchical clustering procedure agree, however, that some of the paintings from the later period are similar to the Renaissance paintings and that at least one of the Renaissance paintings is similar to the paintings from the later period. Let us therefore combine the results of the two procedures into a single "picture" in order to see what we can learn about specific similarities/dissimilarities among the eighteen paintings in our sample. The spatial configuration derived through the multidimensional scaling (Figure 1) and the grouping derived from the hierarchical clustering (Figure 3) are thus combined in Figure 4 (p. 82). Each point has been identified with the appropriate title of the painting.

The circles enclosing clusters of paintings derive from the hierarchical clustering patterns shown in Figure 3 (p. 81). Circles enclose pairs and threesomes and one foursome.

What is immediately striking about this integrated picture is, first, the way in which most of the earlier paintings are bunched together in two clusters in one quadrant; second, the way in which some of the later paintings are quite close to this cluster of early paintings; third, the way in which the rest of the later paintings are widely scattered. In Humanities Working Paper 75, on the basis of the analysis made there, we proposed that the evidence supported the proposition that, whereas there is an identifiable Renaissance school, the later paintings are best evaluated, not as another, distinct "Mannerist" school but as deviations from the Renaissance school. The results of the new analyses of the data reported in this paper are consistent with this proposition.

But we can go beyond this general hypothesis and point out some interesting relationships among specific paintings that the analysis discloses. First, it will be seen that the Renaissance cluster has an inner core consisting of three Raphaels -- the Castiglione, the Belle Jardiniere, and the Angelo Doni.

Let us integrate the analysis we have just made with the analysis presented in Working Paper 75, where, instead of combining the medians for the four dimensions into one composite measure, we presented them separately. It will be seen that the profile for the Castiglione is an almost perfect fit with the Renaissance profile (see Table 5, pp. 75-76).

There might thus be some justification for calling Raphael's Castiglione the Renaissance painting, the Renaissance painting par excellence. However that may be, this finding -- that at the core of the Renaissance school there is a group of Raphaels with the Castiglione at its center -- accords well with art-historical opinion. And recall that the reported finding is based on ratings made by naive raters who not only knew nothing of "schools" and "styles," but most of whom had never so much as seen reproductions of the paintings before they were asked to rate them. In a word, since our raters' view of the paintings was uncontaminated by art historical "theory," it can be said that their ratings, reached independently, tend to support that theory.

Next it should be noted that, surrounding this inner core are other groups, each comprised of a pair of paintings -- the Eleanor of Toledo and another Raphael; the Mona Lisa and Titian's Charles V, and, a bit more remote from the central core, Bronzino's Holy Family and Albertinelli's Noli Me Tangere. Little of this will cause surprise --

even the most ardent advocate of Mannerism is unlikely to think of the Charles V as "mannerist," and most art historians will see a good deal of stylistic difference between that painting and the other Titian in our set -- the portrait of Pope Paul III and his nephews -- just as our raters do.

But a word needs to be said about the two Bronzinos in this cluster, and especially about the Eleanor, since art historians who identify a Mannerist style are likely to regard Bronzino as a member of this school. Have our raters gone astray here? We think not. If one turns back again to our analysis in Working Paper 75, where the medians for the four dimensions were analyzed separately, it will be seen that on three of the dimensions this painting was rated as even more strongly Renaissance than the Renaissance profile (see Table 6, pp. 77-78). It is only on D-1 that our student-raters gave it a non-Renaissance rating, but this was an almost unanimous Z (the most un-Renaissance) rating.

The contrast between the profile of the Eleanor and that of the Bartolomeo Panciatichi (Table 6) is striking. Our raters perceived the later painting as deviating markedly from the Renaissance profile on all four dimensions. That is, they perceived it -- in the language of art historians -- as much more manneristic, and here again we believe most art historians would agree. For our part, we are struck by the differences among Bronzino's oeuvres; perhaps art historians who regard Bronzino as a typical Mannerist are fixing their attention on paintings like the B. Panciatichi (precisely because these paintings do deviate from the Renaissance norm) and overlooking the extent to which other paintings by Bronzino conform to that norm. They may be looking at paintings like the Eleanor through a lens that has been refracted to fit the B. Panciatichi.

This is easy for laymen like ourselves to say, and it will carry no weight unless and until it is supported by expert opinion. But we venture to hope that our findings -- or rather the observations of our naive raters -- will persuade experts to look again, by suggesting to them that the raters have seen similarities and dissimilarities that the experts themselves may have overlooked. One example: The hierarchical clustering procedure shows a very close similarity between the Eleanor (Table 6, pp. 77-78) and the Maddelana Doni (Table 5, pp. 75-76). Since these two paintings are usually discussed in different sections or chapters of art historical writings we suspect that few people, including art historians, look at them together, but we believe that if they are put side by side a strong family resemblance will be seen. It is important in this connection to recall that our subjects did not see the paintings side by side -- as a matter of fact, they were rated by different groups of subjects. However, since these different groups used the same rating scales similarities emerged which can now be confirmed by putting the paintings side by side.

This brings us to the Albertinelli Visitation, which, somewhat surprisingly in view of its early date, is clustered by the hierarchical clustering procedure with the later group. In contrast, the multidimensional scaling procedure locates it with the Renaissance cluster. Note its proximity to the Angelo Doni; it is about the same distance from the center of the Renaissance core as is the Maddelana Doni (i.e., it is rated as being as similar to the core as that painting is) and it is a bit nearer the center (i.e., more similar) than the Albertinelli Noli Me Tangere. The multidimensional scaling procedure seems to us to give a better view of this painting than the hierarchical clustering procedure, and we think most art historians will agree.

This completes our comments on the Renaissance cluster. We turn now to the remaining paintings, which form what may be called a scattered noncluster. All of these dispersed paintings save one are from the later (after 1525) period, a fact that supports the hypothesis that it is the later paintings that deviate more from the Renaissance norm. The one exception -- the one early painting in this group -- is the Michelangelo Holy Family, which is far from the Renaissance core (i.e., rated as very dissimilar) and which clusters with the Paul III (this pair is very similar) and with the Miracle of St. Mark.

Our raters' view of Michelangelo's Holy Family accords well with art-historical opinion: most art historians who regard Mannerism as a distinctive style regard Michelangelo as the proty-typical Mannerist -- some, indeed, would hold that Mannerism is but the product of the profound influence of the sistine ceiling on Michelangelo's contemporaries. Our raters have perceived in this early work, which pre-dates the ceiling and which dates from the time Raphael was painting the two Doni portraits, features that make it much more similar to paintings like (say) the Rosso Moses than like (say) the Doni portraits. The deviation of the Michelangelo from the Renaissance profile comes out very clearly if its profile (see Table 4) is compared with the profile for the Castiglione.

To sum up these comments on what can be learned from the integrated "picture" (Figure 4): We think we have demonstrated that there is a good fit between art-historical opinion and the integrated results of the multidimensional scaling procedure and the hierarchical clustering procedure. A likely retort may be, "So what? If that is what you've done, what is all the shouting about? You are not telling art historians anything new." The short reply to this is that there is a difference between mere opinions, even the opinions of experts, and judgments based on public procedures. We will spell out this short answer in a little detail.

First, and so far as this paper alone goes, if there were no fit, or very little fit, between art-historical opinion and the results of our scaling procedures, we could conclude that our scales had been badly chosen, not that art-historical opinion was mistaken. But the fact that naive raters using our four scales made judgments that correspond so closely to art-historical opinion supports the construct validity of these scales. They were well, or at least serendipitously, chosen. Given this fact, our results are interesting, and perhaps important, precisely because they do not fit perfectly with art-historical opinion. The fit is good enough, we think, to lead one to ask oneself why the fit is not complete. It is possible, of course, that the failures to fit are all due to mistakes by our raters; but it is also possible that in some cases our naive raters are providing us with a "corrected" view of the paintings -- corrected in the same way that the vision of near-sighted and far-sighted people can be corrected to 20/20 -- i.e., "standard" -- seeing. Thus, whereas different art historians view the paintings through different lenses, each of which is a culturally determined "artificial construct," our raters were viewing the paintings through the same lens, viz. the definitions of the dimensions which we gave them.

It is true that though all our viewers used the same lens, we are not yet in a position to say that this is a standard lens, in the way that the occulist's correction to 20/20 is standard. But we can say that, to the extent that our dimensions prove useful in more investigations -- in the art-historical field and in other humanistic disciplines -- they will increasingly become standard, that is, become reliable corrections of individual subjective lenses. This being the case, we believe that experts should not reject out of hand ratings that are at variance with their own perceptions of the paintings.

Second, where art-historical opinion is divided, as over the question of Mannerism, our procedures make it possible to terminate the disagreement by measurement or at least to convert it into a semantical dispute. Our procedures are a device which translates "look alike" and "look different" into measurable distances on a spatial model. As measured on our scales, the Raphaels, the Mona Lisa and the Charles V are much more alike than are the Pontormo, the Salviati, the Tintoretto and the Rosso: We have measured the dispersion of the latter group and compared it with the concentration of the former.

As we wrote in Working Paper 75, the procedure we used there was equivalent to introducing temperature readings into a dispute about whether it was hotter in Los Angeles or in Pasadena on such-and-such a day -- a dispute that would remain nonterminable as long as people had only their subjective feelings to go on. Now, as a result of the more sophisticated procedures used in this paper, we can do even better, and so expand our analogy. Instead of providing only a measure of temperature, we now provide measures of humidity, air quality and wind velocity as well, and a method to organize degrees of similarity-dissimilarity based on all these four features. Suppose, then, armed with these resources, we enter an argument about whether weather conditions in the region around Los Angeles and weather conditions in the region around Memphis form two distinct families or only one family. Using the clustering methods employed in this paper we can say, "Well, the variations from one point to another in the Los Angeles region are such-and-such, and the variations from one point to another in the Memphis region are so-and-so. Do you call that the same, or do you call that different?" So, if someone wants to say that paintings as greatly dispersed as are the Pontormo, the Salviati, the Tintoretto, and the Rosso are all members of the same school, we will not gainsay him.

But is our method reductive? This is like asking whether it is reductive to translate "hot enough to fry an egg" and "cold enough to freeze hell over" into numbers that correspond to the height of a column of mercury in a glass tube. It is impossible, we think, to give a general answer to questions of this kind: Everything depends on whether one is interested in what one is left with after the data have been quantified. Something is lost in every quantification: What is lost in translating "hot enough to fry an egg" into (say) "98^{o}" is expressive power; what is gained is precision: A procedure which can measure weather conditions in Memphis and weather conditions in Los Angeles on a comparable scale will be irrelevant for those who are interested in expressive power; the reduction will be a "bad" reduction. For those who are interested in resolving unnecessary disagreements the reduction will be a "good" reduction. So with the reduction involved in our procedures with the paintings.

To distinguish as we have done, between good and bad reductions and to relativize them to the varying interests and "needs" of different readers seems to us a sensible way of looking at the matter. But it will not please those humanists for whom "reduction" is a powerful pejorative, enabling them to ignore -- to write off -- whatever they dislike. These humanists will certainly resist our attempt to neutralize the term -- our attempt to reduce "reduction." But this should not discourage us. We have known from the outset that, though we can lead humanists to the bridge we are suspending over Lord Snow's chasm, we cannot force them to cross it with us.

Finally, lest it be thought that we regard our procedure as a be all and end all, a few caveats are in order. Caution is necessary in generalizing our analysis of the multidimensional scaling and the hierarchical clustering. First, multidimensional scaling provides a solution in which the total configuration, including the relative distance between any pair of paintings, is affected by all the other paintings. As new paintings are added to the multidimensional analysis the locations of all of the already plotted paintings may change. Hence, to take a possible case: If our raters had rated some seventeenth century paintings along with the sixteenth century works they did rate, the integrated picture of Figure 4 might look very different, in that all of the sixteenth century paintings might form a tight, relatively undifferentiated cluster as compared with the seventeenth century paintings. Second, the results are also dependent on the scales and the rating procedures which were used as the basis for similarity, on the procedure for combing the scale values to derive a similarity value and on the restrictions imposed by the analytical models. That the spatial model developed from the multidimensional scaling procedure shows an excellent fit to the similarity-dissimilarity between data we obtained for the paintings is not surprising. Only four scales were used to develop the Euclidean measure of similarity, and a four dimensional model should capture the structure of such similarity-dissimilarity. It is possible to predict that a two dimensional model such as the one presented here is likely to have little stress.

But our purpose in this paper was not to demonstrate that the spatial model we have used fits the "reality" of paintings better than other possible models, such as taxonomies or hierarchical clusters. Rather, we wish to present a "package" -- a set of procedures which are available for use in many different kinds of humanistic inquiries. We believe we have shown that the package works on a known case. This demonstration should increase confidence that the approach can be used to analyze problems where the outcome is in doubt because a genuinely new hypothesis is being seriously tested, one that could be refuted by the results.

NOTES

1. Kendall Tau nonparametric correlations were also computed and the values are similar in pattern.

2. Similarity-dissimilarity values were also computed using the absolute differences between ratings. This procedure is similar to that used, except that the differences between paintings are not squared and the absolute value of these differences is summed. The outcomes of multidimensional scaling and hierarchical clustering are much the same except that the gamma values are a little lower.

3. These seeming anomalies are examined below in connection with our discussion of individual paintings.

REFERENCES

Chang, J.J. and Carroll, J.D. 1968. "How to Use PROFIT, a Computer Program for Properly Fitting by Optimizing Nonlinear or Linear Correlation. Bell Laboratories (unpublished).

D'Andrade, R.G. 1978. "U-Statistic Hierarchical Clustering." Psychometrika, 43 (1):59-67.

Faust, M.S., Faust, W.L., Jones, M.M. and Jones, W.T. 1980. "Nonterminating Disagreements and Implicit Presuppositions: B.F. Skinner and Carl R. Rogers," Social Science Working Paper No. 357, California Institute of Technology, Pasadena, California.

Freeman, L.C. 1965. Elementary Applied Statistics: For Students in the Behavioral Sciences. New York: John Wiley and Sons (gamma p. 79-87).

Hubert, L.J. and Schultz, J. 1976. "Quadratic Assignment as a General Data Analysis Strategy." British Journal of Mathematical and Statistical Psychology, 29:190-241.

Jones, W.T. 1970. "Philosophical Disagreements and World Views." Proceedings and Addresses of the American Philosophical Association, 43.

__________. 1972. "World Views: Their Nature and Their Function." Current Anthropology, 13 (1).

__________. 1975. The Romantic Syndrome, second edition, with supplementary essay, Nijhof, The Hague.

__________. 1976. "World View and Asian Medical Systems." in Towards a Contemporary Study of Asian Medical Systems, edited by C. Leslie, Berkeley and Los Angeles, CA: University of California Press.

Jones, W.T., Faust, W.L., Faust, M.S. and Jones, M.M. 1980a. "Some Implicit Presuppositions in the Disagreement Over the DNA Guidelines," Social Science Working Paper No. 354, California Institute of Technology, Pasadena, California.

__________. 1980b. "Some Implicit Presuppositions of Typical Writings in the Field of American Intellectual History," Social Sicence Working Paper No. 355, California Institute of Technology, Pasadena, California.

__________. 1981. "Paintings and Their Implicit Presuppositions: A Preliminary Report," Humanities Working Paper No. 66, California Institute of Technology, Pasadena, California.

__________. 1982. "Paintings and Their Implicit Presuppositions: High Renaissance and Mannerism," Humanities Working Paper No. 75, California Institute of Technology, Pasadena, California.

Kruskal, J.B., Young, F.W. and Seery, J.B. 1973. "How to Use KYST, a Very Flexible Program to do Multidimensional Scaling and Unfolding." Bell Laboratories (unpublished).

Romnoy, A.K., Kieffer, M. and Klein, R.E. 1973. "A Normalization Procedure for Correcting Biased Response Data." Social Science Research, 2 (4):307-320.

TABLE 1

MEDIANS OF INDIVIDUAL PAINTINGS FROM THE EARLY PERIOD, 1500-1515

	D-1	D-2	D-3	D-4
Albertinelli: Noli Me Tangere	O (6)+	B (2)	O (6)	E (5)
Albertinelli: Visitation	A (1)	C (3)	A (1)	X (9)
Raphael: Angelo Doni	B (2)	O (6)	B (2)	A (1)
Raphael: Maddalena Doni	Y (10)	A (1)	A (1)	A (1)
Raphael: Belle Jardiniere	B (2)	B (2)	A (1)	B (2)
Leonardo: Mona Lisa	O (6)	C (3)	B (2)	A (1)
Michelangelo: Doni Holy Family	A (1)	X (9)	V (7)	V (7)
Raphael: Castiglione	C (3)	C (3)	A (1)	B (2)
Renaissance profile-med of meds	B-C	C	A-B	B

TABLE 2

MEDIANS OF INDIVIDUAL PAINTINGS FROM THE LATER PERIOD, 1545-1560

	D-1	D-2	D-3	D-4
Pontormo: Visitation	X (9)+	O (6)	C (3)	Y (10)
Bronzino: B. Panciatichi	Z (11)	O (6)	X (9)	O (6)
Bronzino: Noli Me Tangere	O (6)	Y (10)	D (4)	Y (10)
Rosso: Moses and the Daughters of Jethro	Z (11)	B (2)	Z (11)	Z (11)
Salviati: Caritas	B (2)	Y (10)	A (1)	Z (11)
Bronzino: Holy Family	O (6)	B (2)	O (6)	O (6)
Bronzino: Eleanor of Toledo	Z (11)	A (1)	A (1)	A (1)
Titian: Charles V	W (8)	B (2)	B (2)	C (3)
Titian: Paul III	B (2)	V (7)	W (8)	V (7)
Tintoretto: Miracle of the Slave	C (3)	Y (10)	X (9)	Z (11)
Profile of late paintings-med of meds	V	O	E	W

+ The alphabetical letter ratings used in Working Paper 75 have been transformed into numbers for the purposes of multidimensional scaling.

TABLE 3

Intercorrelation among the four scales. Correlations are computed on ratings of the paintings.

Scales	2	3	4
1	N.S.	N.S.	N.S.
2		N.S	<u>.61</u>
3			.52

The underlined values are significant $p < .05$.

The other values are probably only different than 0.0 because of chance factors. Indicated by N.S.

TABLE 4

QUADRATIC ASSIGNMENT -- DISTANCES AND Z VALUES

	Average distance	Significance of difference between average distance among all paintings and average distance among Renaissance paintings (and also later paintings)
All paintings	2.64	
Renaissance paintings	2.02	Z = 2.03 (sig.)
Later paintings	2.80	Z = 0.68 (not sig.)

TABLE 5

MEDIAN VALUES FOR EIGHT RENAISSANCE PAINTINGS EACH IN RELATION TO THE PROFILES OF PAINTINGS OF EARLIER AND LATER PERIODS

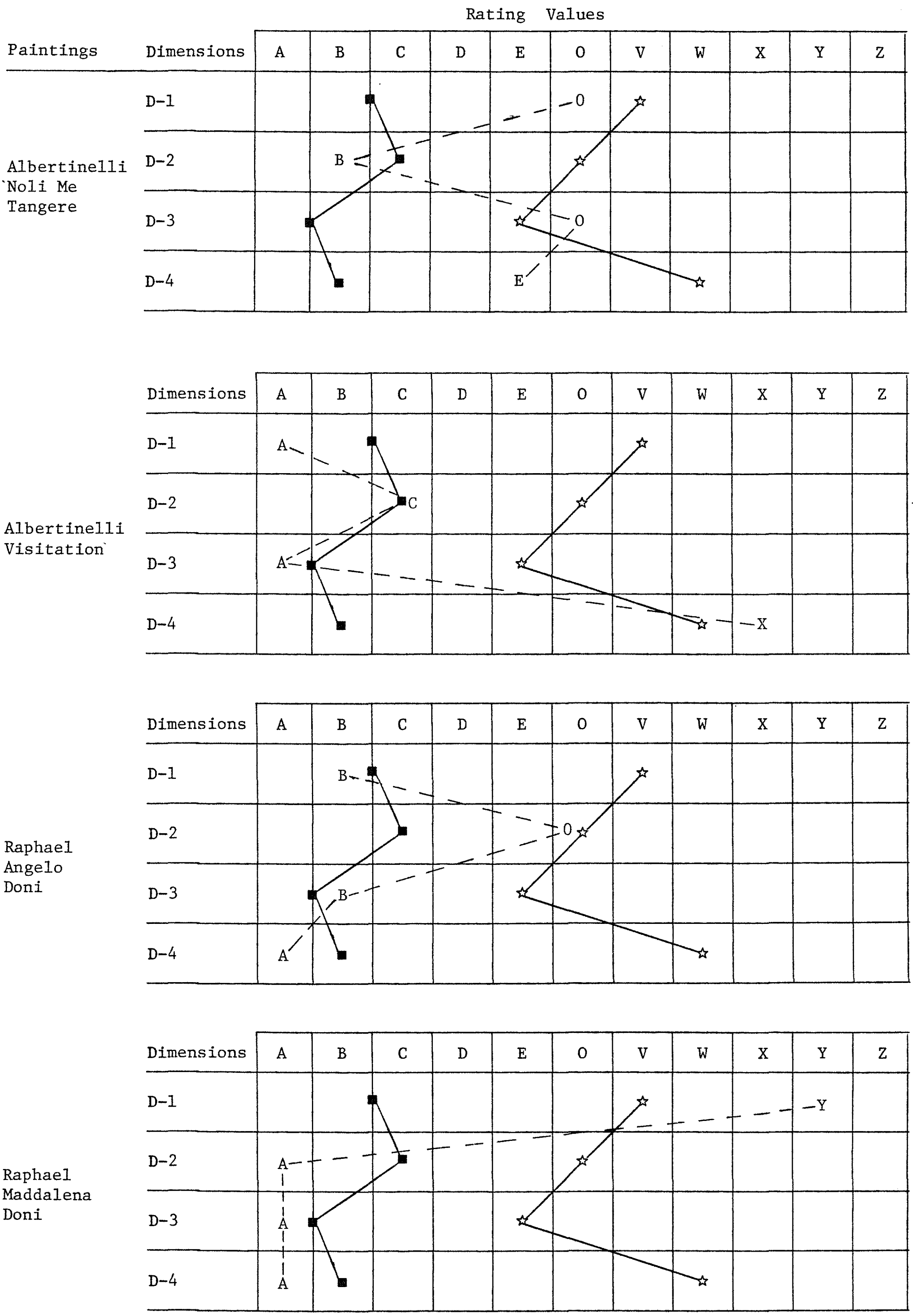
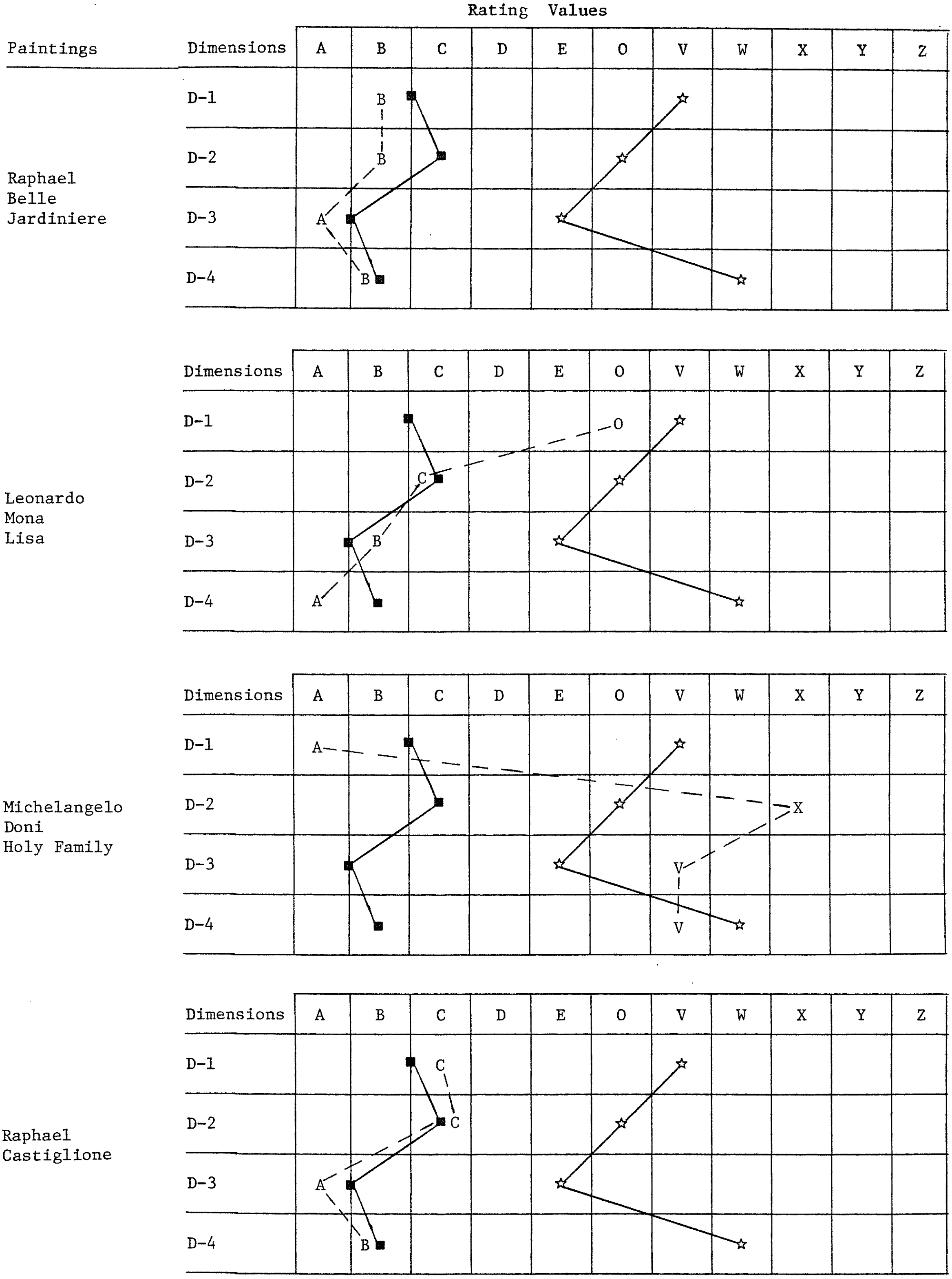


TABLE 5 (cont.)



Key:

■——■ Earlier median

☆——☆ Later median

----- Individual painting

The solid lines connecting the medians for the profiles of Earlier and Later paintings and the dashed line connecting the medians for that painting are for pictorial clarity--intermediate values should not be interpolated.

TABLE 6

MEDIAN VALUES FOR TEN PAINTINGS OF LATER PERIOD EACH IN RELATION TO THE PROFILE OF PAINTINGS OF EARLIER AND LATER PERIODS

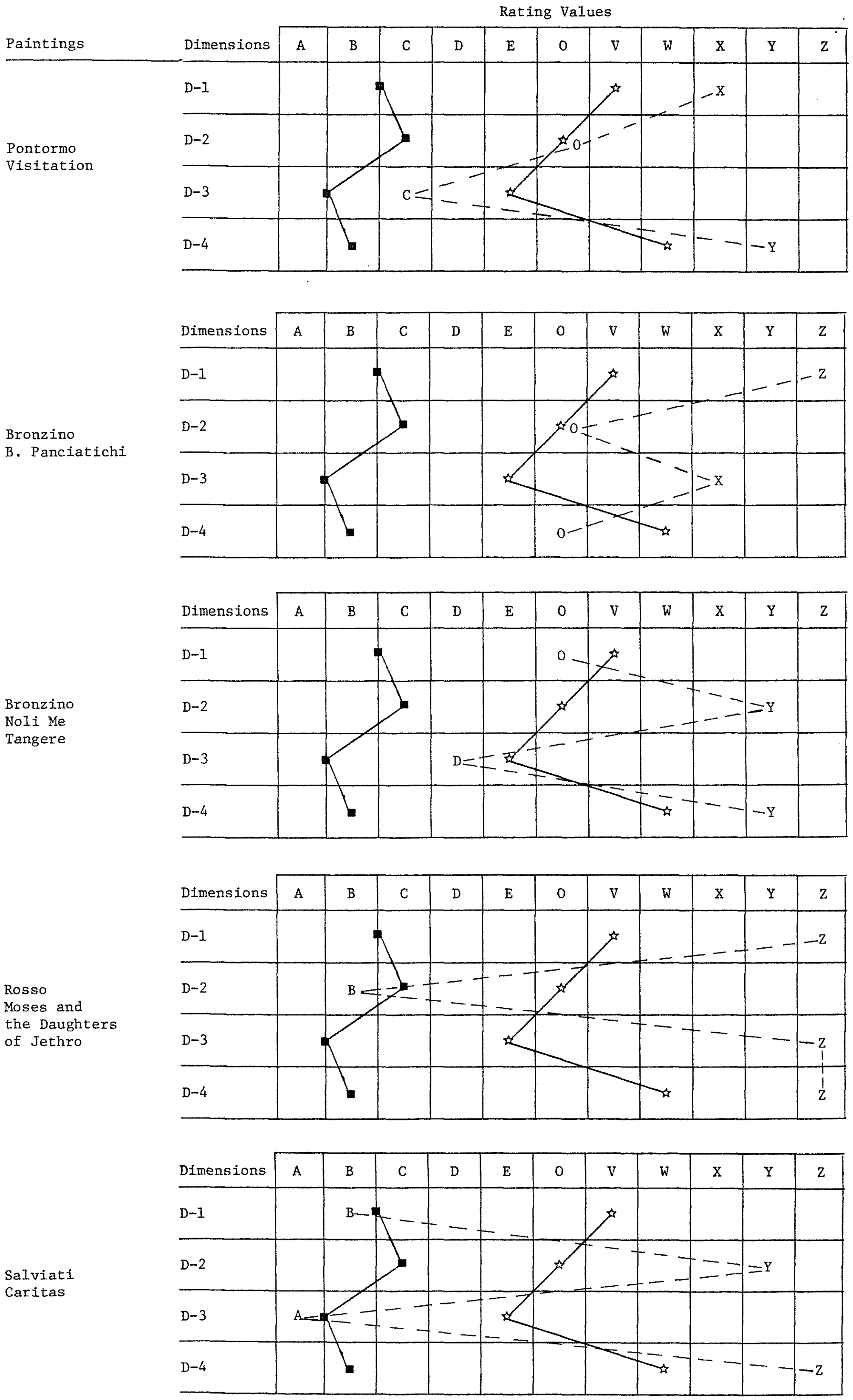
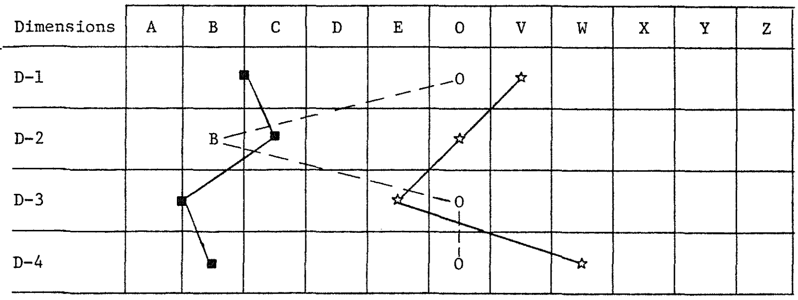


TABLE 6 (cont.)

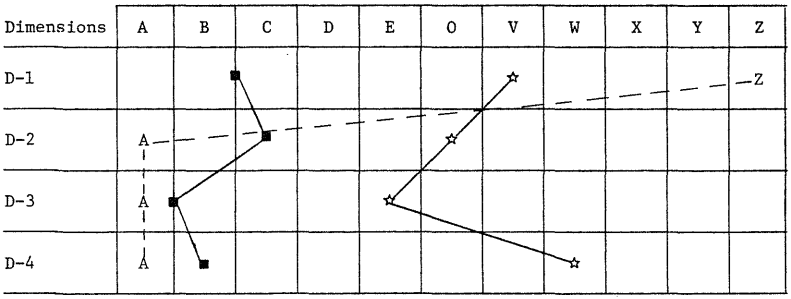
Rating Values

Paintings

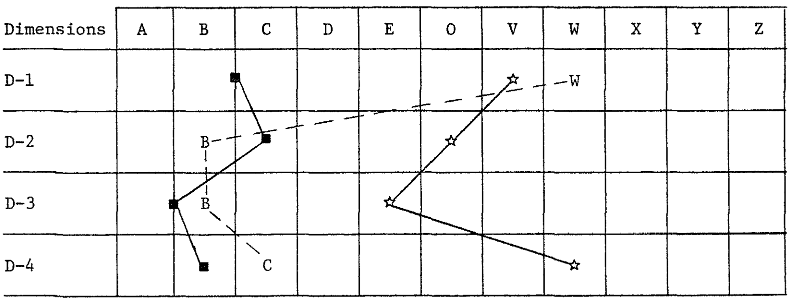
Bronzino
Holy Family



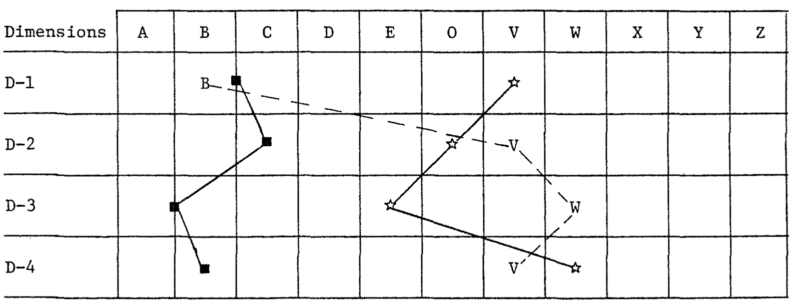
Bronzino
Eleanor of Toledo



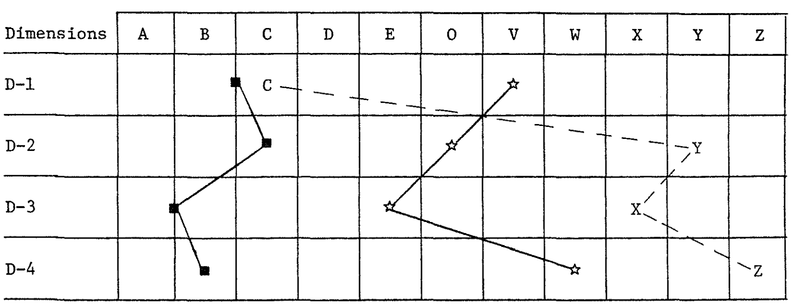
Titian
Charles V



Titian
Paul III



Tintoretto
Miracle of
the Slave



Key:

■——■ Earlier median

☆——☆ Later median

----- Individual painting

The solid lines connecting the medians for the profiles of Earlier and Later paintings and the dashed line connecting the medians for that painting are for pictorial clarity--intermediate values should not be interpolated.

FIGURE 1

CONFIGURATION PLOT IN TWO DIMENSIONS. RENAISSANCE (R) AND LATER (L) PAINTINGS, INPUT IS DISTANCES BETWEEN NORMALIZED SIMILARITY VALUES

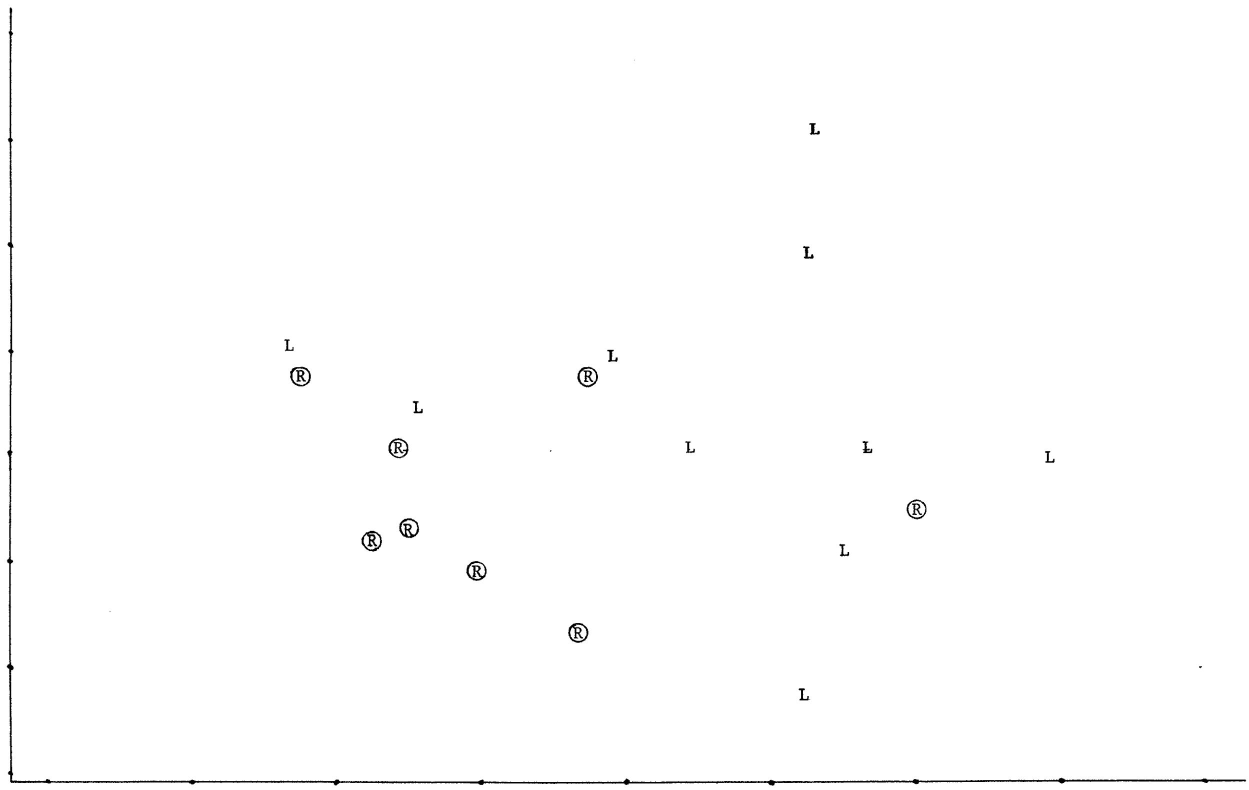


FIGURE 2

CONFIGURATION PLOT IN TWO DIMENSIONS. RENAISSANCE (R) AND LATER (L) PAINTINGS, INPUT IS DISTANCES BETWEEN NORMALIZED SIMILARITY VALUES

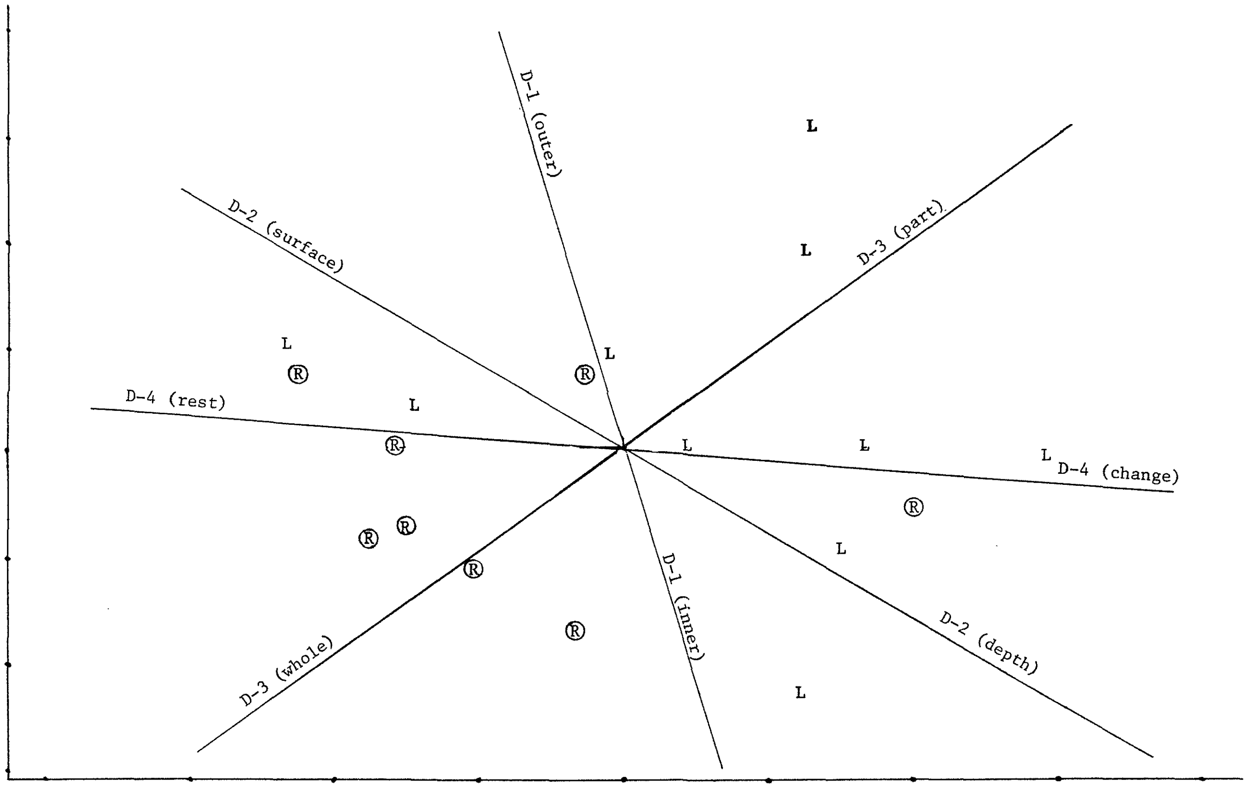


FIGURE 3

HIERARCHICAL CLASSIFICATION

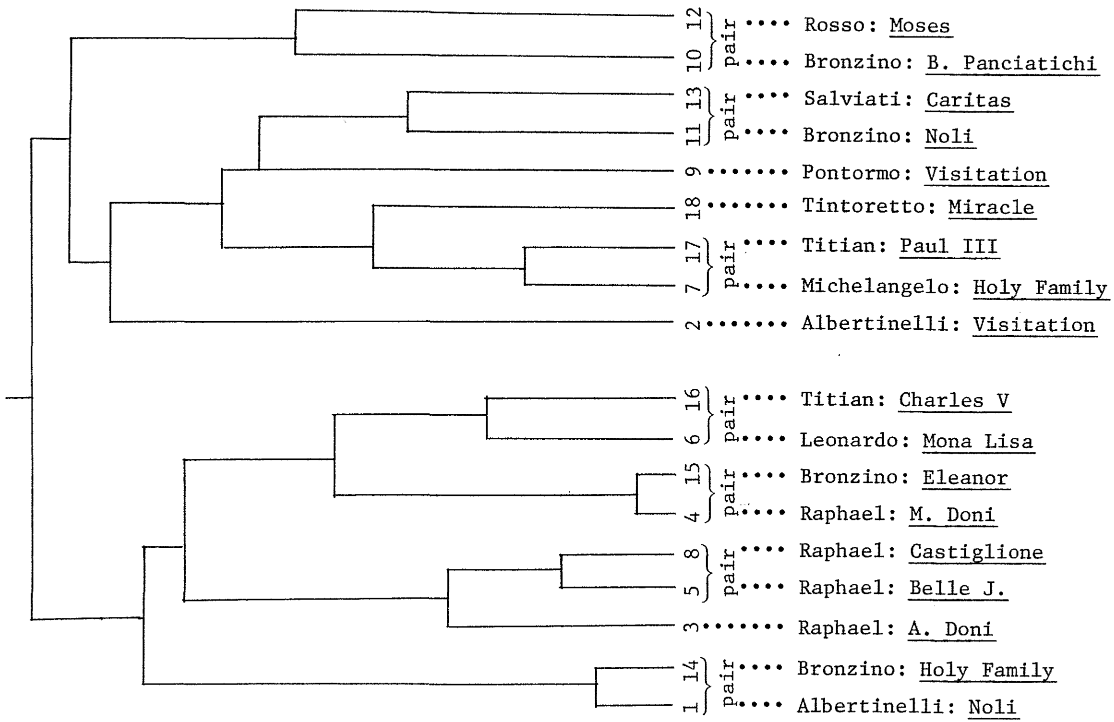


FIGURE 4

CONFIGUREATION PLOT. IN TWO DIMENSIONS RENAISSANCE AND MANNERIST PAINTINGS
INPUT IS DISTANCES BETWEEN NORMALIZED DIMENSIONS