Gerald L. Lohse, Kevin Biolsi, Neff Walker, and Henry H. Rueter

A Classification of **Visual Representations**

Why do we often prefer glancing at a graph to studying a table of numbers? What might be a better graphic than either a graph or table for seeing how a biological process unfolds with time? To begin to answer these kinds of questions we examine the cognitive structure of graphics and report a structural classification of visual representations.

McCormick, DeFanti, and Brown [16] define visualization as "the study of mechanisms in computers and in humans which allow them in concert to perceive, use, and communicate visual information." Thus, visualization includes the study of both image synthesis and image understanding. Given this broad focus, it is not surprising that visualization spans many academic disciplines, scientific fields, and multiple domains of inquiry. However, if visualization is to continue to advance as an interdisciplinary science, it must become more than a grab bag of techniques for displaying data. Our research focuses on classifying visual information. Classification lies at the heart of every scientific field. Classifications structure domains of systematic inquiry and provide concepts for developing theories to identify anomalies and to predict future research needs.

Extant taxonomies of graphs and images can be characterized as either functional or structural. Functional taxonomies focus on the intended use and purpose of the graphic material. For example, consider the functional classification developed by Macdonald-Ross [14]. One of the main categories is *technical diagrams* used for maintaining, operating, and troubleshooting complex equipment. Other examples of functional classifications can be found in Tufte [22]. A functional classification does not reflect the physical structure of images, nor is it intended to correspond to an underlying representation in memory [1].

In contrast, structural categories are well learned and are derived from exemplar learning. They focus on the form of the image rather than its content. Rankin [18] and Bertin [2] developed such structural categories of graphs. Rankin used the number of dimensions and graph forms to determine his classification of graph types. Major categories in this scheme include rectilinear cartesian coordinate graphs, polar coordinate graphs, bar graphs, line graphs, matrix diagrams, trilinear charts, response surfaces, topographic charts, and conversion scales.

We adopt the view that visual representations are data structures for expressing knowledge [11, 19]. As such, visual representations can facilitate problem-solving and discovery by providing an efficient structure for expressing the data. Cognitive efficiency results when perceptual inferences replace arduous cognitive comparisons and computations. Since the primary advantage of visual information is that the representation conveys the data structure directly, we chose to develop a structural classification.

Few previous taxonomies and classification schemes for visual representations are based on experimental data; most rely instead simply on the author's intuitions. While these intuitions have yielded valuable insights, empirical work is required to discover and elaborate the basis on which people organize visual information. Our research focuses on how people classify visual representations into meaningful, hierarchically structured categories.

In Lohse *et al.* [12], we began an exploratory research program aimed at classifying visual information. Our classification was based on subjects' ratings of the visual similarity between visual representations, and we identified six basic categories of visual representations: graphs, tables, maps, diagrams, networks, and icons. In addition, we tentatively identified two dimensions that distinguish these clusters. One dimension suggested that a graphic could express either continuous or discrete information, while the second dimension suggested that some visual representations are more efficient than others for conveying information.

Our second research study [13] systematically examined the effect of graphic arts training on the classification of visual representations. In that study, subjects without graphic arts training classified graphics into the same groups found in [12], while those with graphics arts training classified them into slightly different groups. This difference in classification resulted from the influence of the color, form, pattern, and overall shape of the stimuli on the graphic artists' similarity judgments.

In the current study we confirm the basic categories from these initial investigations and construct a classification of visual representations. Specifically, we identify features that characterize high-level categories of visual representations. For instance, what characteristics distinguish maps from diagrams? The results describe the attributes that people may use to judge similarity among visual representations. Through an understanding of the taxonomic relationships among a broad range of graphics, we hope to help designers decide how to represent various kinds of information, as well understand the limitations of different visual representations for conveying certain types of information.

First, we describe our data collection methods for ratings of graphics on 10 scales and sortings of these graphics. Next, we present the results of our analysis of this data using several multivariate analysis techniques. Then, we discuss the 11 categories of visual representations that emerged from the classification, describing the characteristics of each. After describing some anomalies that have important implications for the design of graphics, we propose directions for future classification research as well as suggestions for the human-computer interface for visualization tools.

Methods

Materials. The 60 graphical items shown in Figure 1 were used in this study. Forty of these items were those used in our earlier work [12, 13]. The original 40 items were selected to be as representative as possible within the domain of static, two-dimensional graphic representations. To insure adequate variation, we consulted popular books on graphics [2, 22]. An additional 20 items expanded under-represented categories of graphics as identified by our multidimensional scaling solution [12] and by comments from reviewers of our previous studies [12, 13].

Subjects. Sixteen subjects were recruited from the students and staff of the University of Michigan. Half of the subjects had a high degree of knowledge about graphic design, as these subjects either had a master's degree in Fine Arts or were currently enrolled in a Master of Fine Arts program. The remaining eight subjects were all currently enrolled graduate students who had no special training in art or graphics. Subjects were paid $20 for participating in the study. (Due to experimenter error, the sorting data for subject 12 was unusable; this subject was removed from all subsequent analyses.)

Procedures. Subjects performed three tasks in a two-hour session. These tasks were naming, rating, and sorting the 60 items. First, subjects examined all 60 items and named each one to insure that they were familiar with the entire range of items before beginning the rating task. This helped reduce the effects of order of presentation or anchoring effects on the subjects' subsequent ratings.

Next, subjects rated each of the 60 items on 10 nine-point Likert scales. The 10 rating scales were derived from a frequency analysis of keywords used by subjects to describe each cluster of items during the sorting task of our two previous studies [12, 13]. Each unique word was tallied, and we collapsed this data across similar word phrases, keywords, or synonyms. The authors selected 10 scale items from the final collapsed list of unique keywords. The 10 scales and their anchor-point phrases were:

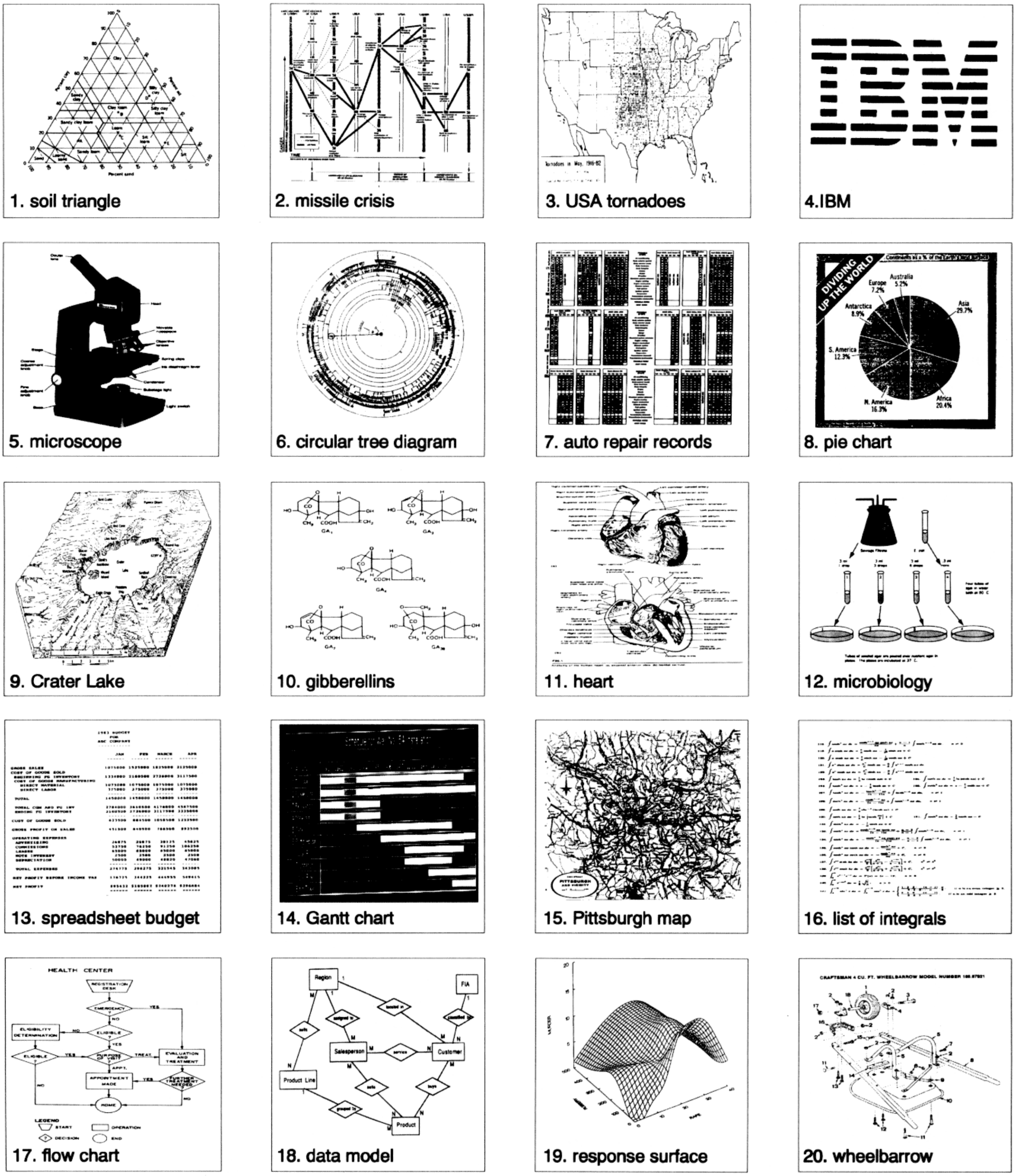
- spatial-nonspatial,
- nontemporal-temporal,
- hard to understand-easy to understand,
- concrete-abstract,
- continuous-discrete,
- attractive-unattractive,
- emphasizes whole-emphasizes parts,
- nonnumeric-numeric,
- static structure-dynamic process,
- conveys a lot of information-conveys little information

Figure 1. Sixty visual representations used in the classification study

The 60 items were presented to the subjects in two books for rating. In the books, each graphic and the 10 nine-point Likert scales were on facing pages. Subjects were asked to rate each item on all 10 scales before turning the page. The order of the 60 items in the booklets was randomized for each subject. Subjects were allowed to take as much time as needed and were allowed to take breaks during the rating procedure.

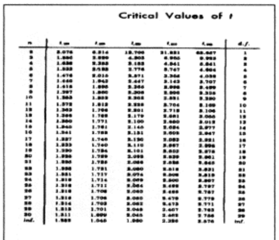
The final procedure was a bottom-up sorting task. For this task, the 60

1A

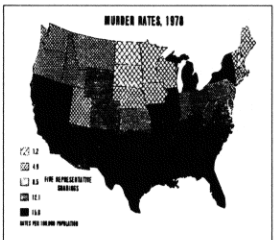


items were placed randomly on a large table, and the subjects were asked to sort them into groups of similar items. Subjects were given no explicit criteria for judging similarity and could create any number of groups and any number of items per group. Once the subjects had completed their initial groupings, they described each group and explained why all the items in the group were similar. After the experimenter recorded these descriptions, the subjects grouped their initial groupings into higher-order clusters of similar groups. Again, the experimenter recorded the subjects' explanations of why all the items within a cluster were similar. This process was repeated until all 60 items were placed in a single group.

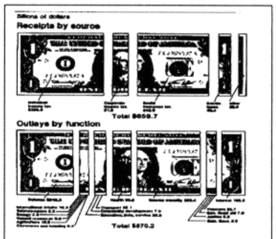
1B



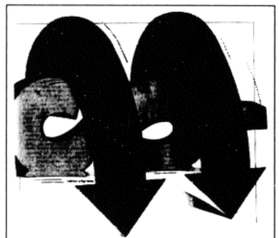
21. t-table

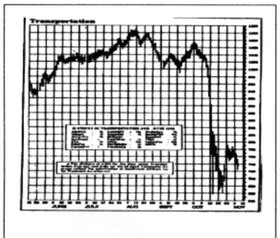


22. USA murder rates

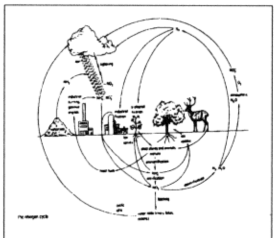


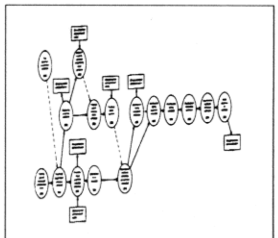
23. dollar bar chart

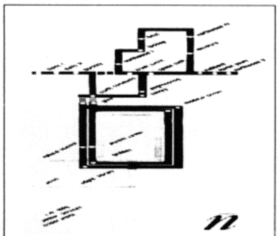

24. yen vs. dollar

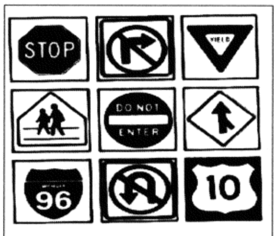

25. Dow Jones averages


26. UM campus

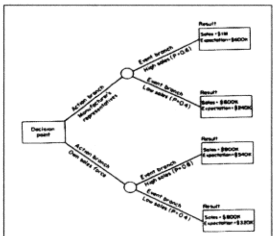

27. nitrogen cycle


28. pert chart

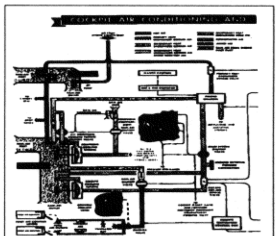

29. bus routes



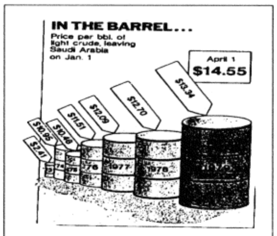
30. highway signs



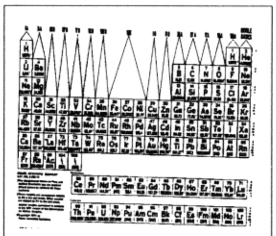
31. decision tree

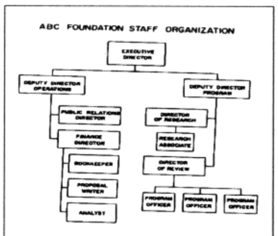

32. cockpit air conditioner


33. Tale of two cities

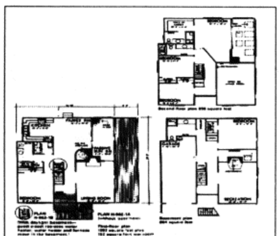


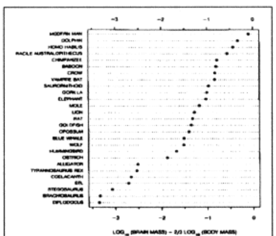
34. oil barrels


35. periodic table

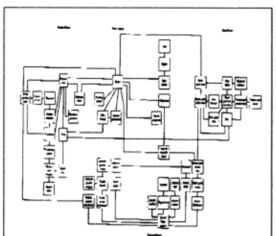


36. organizational chart


37. floorplan


38. dot chart

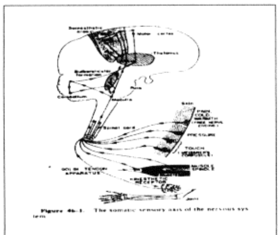

39. music

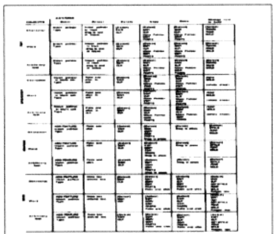

40. conceptual data model

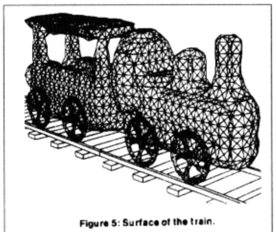
Results

We first clustered the subjects to determine whether any were obvious outliers among the subjects who sorted the graphics. As a measure of similarity between pairs of subject sorts, we used the Jaccard coefficient [9], which is computed as follows: Jaccard(i,j) = A/(N − B), where A is the number of pairs of graphics in which the members of the pair appear in the same groups for the sortings of both subjects i and j; B is the number of pairs of graphics for which the members of the pair appear in different groups in both subjects' sorts; and N is the total number of graphic pairs or n(n−1)/2, where n is the number of graphics.

Complete linkage cluster analysis [10] was then applied to the matrix of

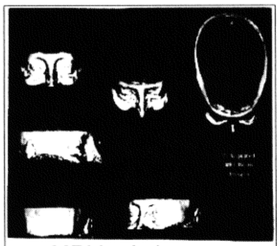

41. nervous system


42. software features

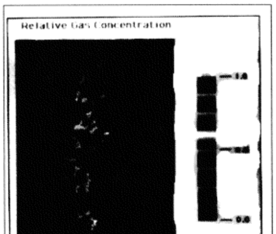


43. surface of train

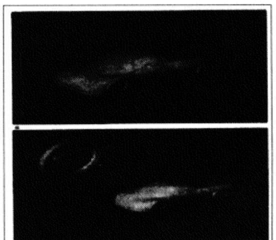

44. city buildings

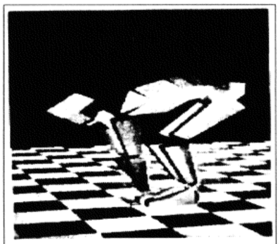

45. MRI brain images

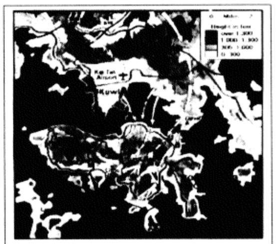

46. world erosion

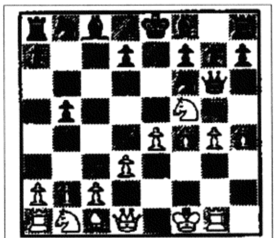


47. gas concentration

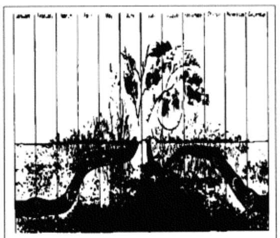

48. CAD image of a car

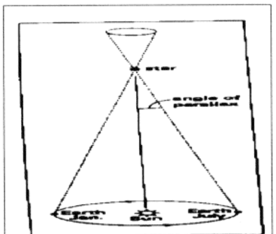

49. robot

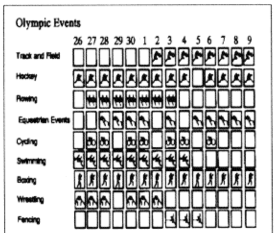

50. map of Hong Kong


51. chess board

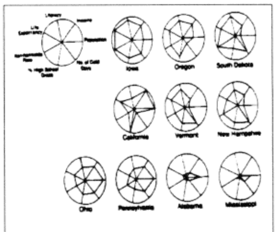

52. fertility rates


53. beetle life cycle

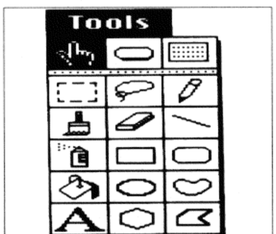

54. equinox cycle



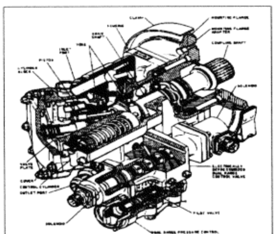
55. Olympics schedule

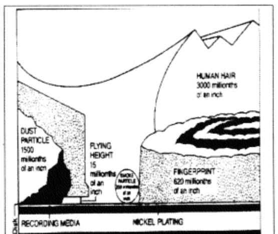

56. star diagrams


57. DNA double helix



58. iconic menu


59. engine block



60. diskette cross section

Jaccard coefficients. The resulting tree diagram suggested that subject 11 sorted the graphic items in a manner different from the other subjects, and therefore, the data for this subject was removed from all subsequent analyses. The sortings of subject 11 were based almost entirely on the subject matter of the graphics. For example, her named categories included "biology," "economics," and "sports." The Jaccard cluster analysis revealed no other outliers; all the remaining subjects based their ratings primarily on the type of knowledge conveyed by the representation (e.g., spatial, discrete, abstract, temporal, etc.) rather than its functional area. This virtual unanimity may have been an artifact of the data collection process. Since subjects rated the items on Likert scales before sorting the items into similar groups, the Likert scale rating task may have biased the subsequent sorting task. However, only the complete linkage cluster analysis results would be affected. All other statistical analyses used the Likert scale data.

In order to identify groups or clusters of items in the subjects' sortings, a matrix of similarities was constructed by counting the number of times each pair of graphics was grouped together in the subjects' lowest level sorts. For example, graphics number 4 and number 30 appeared in the same initial grouping for 11 of the 14 subjects; therefore the corresponding entry in the matrix is 11. The entries in the matrix ranged from 0, when two graphics never appeared together, to 14, when the graphics appeared together in every subject's lowest level sort. The similarity matrix was then used as the basis for complete linkage hierarchical clustering. The resulting tree had nine primary classes or clusters of graphics,[1] two of which had subclasses. These 11 classes are described.

[1]Because of the large number of ties in the 60 × 60 proximity matrix, complete link clustering can produce different solutions depending on the order in which entries of the input similarity matrix are considered. Therefore, we performed six complete link clusterings on six different random orderings of the similarity matrix. From these six clustering solutions we derived the classes shown in Table 1 and the three singleton classes corresponding to graphics no. 6, no. 47, and no. 51.

We next sought to determine if the rating scales (that were based on our previous work [12, 13]) were predictive of class membership or clusters derived from the sorting data. To do this, we followed a three-step procedure. First, we used principle components analysis to determine if the 10 scales could be reduced to a smaller set of underlying dimensions. Then we used two different methods for classification: classification trees [3] and discriminant analysis [7]. These techniques were used on the average ratings of the 10 scales for each of the 57 graphical items (i.e., with the three singleton items removed). The items with the highest and lowest rating for each scale are shown in Figure 2.

A principle components analysis of the data revealed that only one scale, amount of information conveyed, explained less than 9% of the total variance (see Table 1). No single scale explained more than 16% of the total variance. The analysis suggests the 10 scales are relatively independent (i.e., nonredundant) and of approximately equal importance (in terms of variance explanation), so we therefore make use of all 10 in the analyses.

The Classification and Regression Trees (CART) methodology [3] was next used to construct a binary classification tree (Figure 3) in order to determine if the ratings on the 10 scales were predictive of membership in the clusters yielded by the hierarchical clustering analysis. Each terminal node of such a tree is associated with a single graphic class (although a single class may label more than one terminal node). The simplest type of classification tree is one in which each internal node of the tree corresponds to a single independent variable, called the splitting variable for that node. Associated with the splitting variable is a threshold value, which determines whether a to-be-classified item is sent left or right in the subsequent branching. Items are classified by running them down the tree and sending them right or left at each node depending on whether or not they exceed the threshold value for the corresponding splitting variable at that node. When an item reaches a terminal node, it is assigned to the class associated with that node.

In order to apply the CART program, ratings on the 10 scales were averaged across all 14 subjects. These average ratings were then used to predict membership for the 11 classes. The singleton classes were excluded from the analysis.

The analysis yielded the classification tree[2] shown in Figure 3. This tree correctly classifies 48 of 57, or 84%, of the graphics. The *true* classification rate is estimated at 53% using a cross-validation estimate [3]. This result strongly suggests the ratings on the 10 scales can be used to predict group membership derived from the sorting data. Table 2 shows the mean Likert scale scores for each of the 11 groups found in the CART analysis.

As an additional check on the relationship between the rating scales and the classes derived from the sorting task, we used discriminant analysis. As with the CART methodology, the purpose of the discriminant analysis was to determine the relationship between the rating scales and the sort-derived classes. Discriminant analysis is a technique whereby linear combinations of a set of independent variables are constructed so as to maximally discriminate among the classes of interest. Various interpretation techniques are then used to determine the discriminability of the classes as well as the contribution of the different variables to the discrimination. Determining the extent to which the variables are effective in predicting class membership is also an integral part of the analysis.

Discriminant analysis was applied to the same data used in the CART analysis. In order to provide clearer interpretation, the resulting discriminant functions were first rotated. Based on both the rotated coefficients for the discriminant functions and the discriminant loadings (i.e., the correlations between the scales and the discriminant functions), the first three functions correspond most strongly to the numeric, information, and spatial scales, respectively. Note that these are the same scales that comprise the first three splitting vari-

[2]For those familiar with the CART methodology, the tree was constructed using the Gini splitting criterion with prior probabilities proportional to the class sizes in the data set and 11-fold cross-validation.

2A

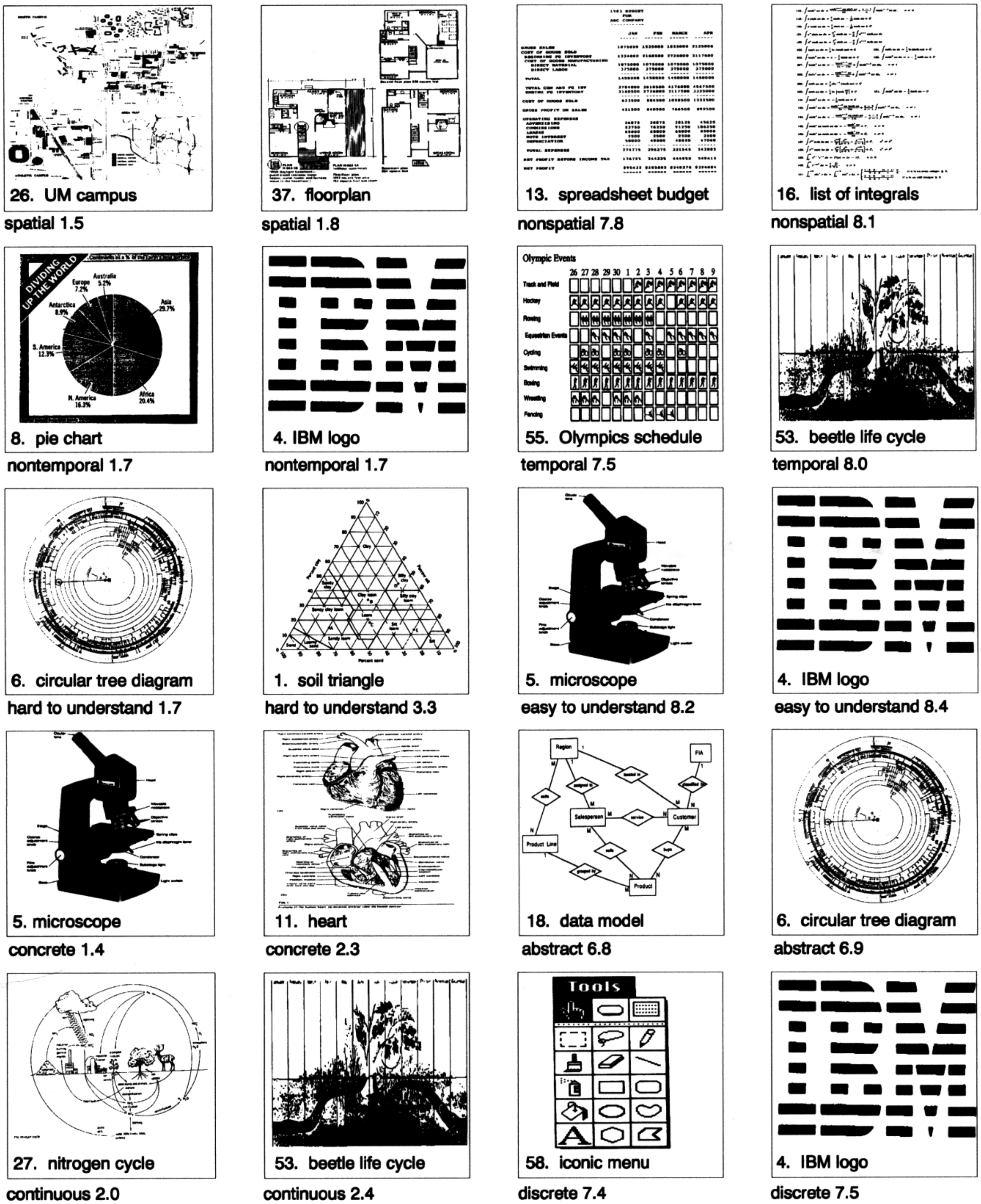
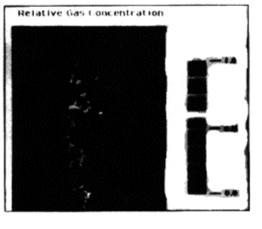


Figure 2. The lowest and highest Likert scale rating for each of the 10 scales used in the classification study

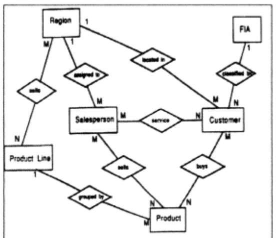


47. gas concentration

attractive 1.5

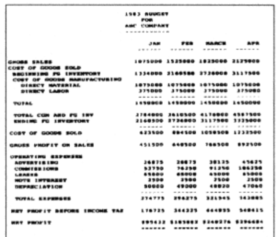
44. city buildings

attractive 1.8



18. data model

unattractive 6.4



13. spreadsheet budget

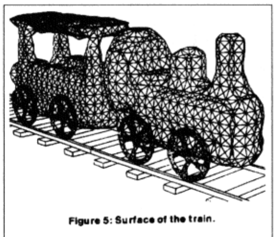
unattractive 6.5

58. iconic menu

emphasizes parts 2.0

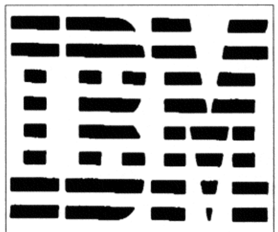
42. software features

emphasizes parts 2.1



43. surface of train

emphasizes whole 7.7

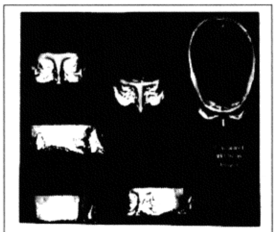


4. IBM logo

emphasizes whole 8.2

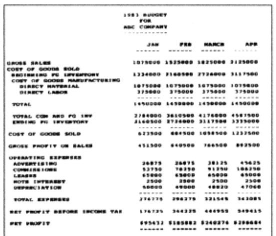
44. city buildings

non-numeric 1.1



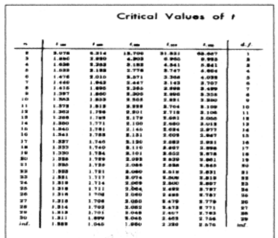
45. MRI brain images

non-numeric 1.3



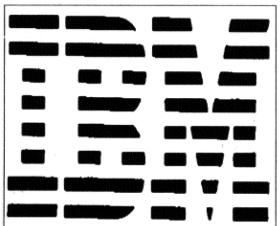
13. spreadsheet budget

numeric 8.7



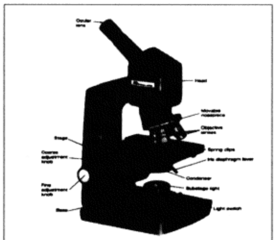
21. t-table

numeric 8.9



4. IBM logo

static 1.2

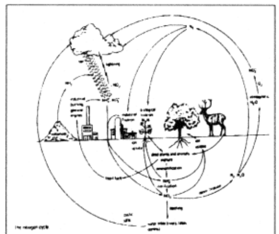


5. microscope

static 1.9

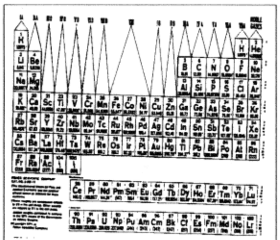
53. beetle life cycle

dynamic 7.8



27. nitrogen cycle

dynamic 7.9



35. periodic table

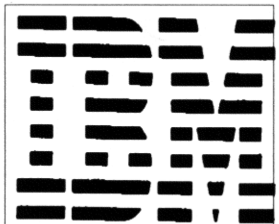
conveys a lot 1.7

15. Pittsburgh map

conveys a lot 2.1

49. robot

conveys a little 7.3



4. IBM logo

conveys a little 7.7

ables in the classification tree.

The discriminant analysis correctly classified 47 of 57, or 83%, of the graphics. The true classification rate is estimated at 56%. These classification rates are nearly identical to those found using the CART methodology. These results, when combined with those of the CART analysis, provide confirmation for our belief that the rating scales are predictive of class membership in sorting and may have served as the basis on which subjects made their sorting decisions.

Overall, our analyses all provide confirmatory evidence of the taxonomic structure of the graphic items presented in Figure 3. In addition, the results of both the CART analysis and the discriminant analysis suggest that the 10 rating scales can be used as predictors of class membership in the classification.

Discussion

High-Level Descriptions of Major Taxonomic Groups

Eleven categories of visual representations emerged from the classification: graphs, tables, graphical tables, time charts, networks, structure diagrams, process diagrams, maps, cartograms, icons, and pictures. Here we describe these major groups and the type of knowledge conveyed by each class of representation.

Graphs encode quantitative information using position and magnitude of geometric objects. One-, two-, or three-dimensional numerical data is plotted on a Cartesian coordinate or polar coordinate system. Common graph types include scatterplot, categorical, line, stacked bar, bar, pie, box, fan, response surface, histogram, star, polar coordinate, and Chernoff face graphs. Except for the star graph (number 56) and the gas concentration graph (number 47), all graphs formed a single cluster. The gas concentration graph was a singleton. The novel format of the gas concentration graph lacked polar or rectilinear axes normally associated with graphs. This may have prevented subjects from grouping the gas concentration graph with other graphs. Star charts resemble pie charts, but subjects did not cluster these items in one group. Subjects grouped star charts with cartograms perhaps because both convey information from a spatial pattern created by the data. Graphs emphasize the whole display as compared with tabular data that emphasize parts of the display.

Tables are an arrangement of words, numbers, signs, or combinations of them to exhibit a set of facts or relationships in a compact format. Tables have less abstract symbolic notation than graphs. For example, tables usually do not contain a legend for mapping symbolic information to semantic information, whereas graphs often contain a legend to accomplish this task. Two groups of tables appeared in the classification: graphical and numerical. The primary distinction depended on how numeric information is coded in the table. Graphical tables, like the auto repair records (number 7), used shading to encode frequency of repair data, whereas the statistical table of the critical values of the t statistic (number 21) shows only numeric data. Numerical tables emphasize parts of the whole representation (e.g., individual data values). Subjects believed numerical tables conveyed a lot of information in an unattractive format but that graphical tables conveyed a lot of information in an attractive format.

Time charts display temporal data. They differ from tables in their emphasis on temporal data. Examples include the Gantt chart (number 14) and the time schedule of Olympic events (number 55). Both examples include graphical objects. The Gantt chart includes bars to indicate the length of an event; the Olympic schedule uses icons to identify each type of event.

Network charts show the relationships among components. Symbols indicate the presence or absence of components. Correspondences among the components are shown by lines, arrows, proximity, similarity, or containment. The planar coordinate system of network charts is generally void of meaning. Meaning results from an efficient spatial arrangement of the data that is parsimonious and avoids intersecting lines. Examples include flow charts, organizational charts, decision trees, pert charts, and data models. Subjects believed network charts conveyed a lot of information.

There are two types of diagrams, both of which express spatial data. **Structure diagrams** are a static description of a physical object. The spatial data expresses the true coordi-

Table 1. Principal components analysis showing percentage of the total explained variance for each of the ten scales used in the ratings task

Scale number	Percent variance	Scale description
1	11.3	spatial-nonspatial
2	10.1	nontemporal-temporal
3	9.6	hard to understand-easy to understand
4	9.9	concrete-abstract
5	10.5	continuous-discrete
6	10.3	attractive-unattractive
7	16.0	emphasizes whole-emphasizes parts
8	9.5	nonnumeric-numeric
9	10.6	static structure-dynamic process
10	2.2	conveys a lot of information-conveys little information

nate dimensions of the object. Examples include the cross-sectional view of an engine (number 59) and a heart (number 11). **Process diagrams** describe the interrelationships and processes associated with physical objects. The spatial data expresses dynamic, continuous, or temporal relationships among the objects in process diagrams. Examples include the nervous system (number 41) and the nitrogen cycle (number 27). While the structure diagram formed a single cluster, the process diagrams were grouped in three clusters. The beetle life cycle (number 53) and the cockpit air conditioner flow diagram (number 32) were grouped with cartograms. The nervous system (number 41) was grouped with structure diagrams. This seems surprising given the prominent visual arrows that show connection and flow; however, all diagram components appear in their true physical locations, an important component of structure diagrams. Structure diagrams convey a lot of spatial, nonnumeric, concrete information. The main difference among similarity measures for maps, cartograms, and structure diagrams is that maps and cartograms express more numeric information than structure diagrams. Thus, people might reason about qualitative relationships from structure diagrams, but reason about qualitative and quantitative relationships from maps and cartograms.

Maps are symbolic representations of physical geography. Maps depict geographic locations of particular features using symbols or lettering. Examples include marine charts, highway maps, topographic maps, land use maps, and various projections of world maps. Maps differ from cartograms in that cartograms superimpose quantitative data over a base map. Therefore, it is not surprising that subjects felt cartograms were more difficult to understand than true maps.

Cartograms are spatial maps that show quantitative data. Examples include chloropleths, isopleths, dot maps, and flow maps. Chloropleths use color, gray scale, or texture to code areas of equal value. Isopleths use lines to join points with the same quantity or value (e.g., contour maps). Dot maps use points or symbols to show the location of individual points on a map. Flow maps show direction of movement by the number, width, and direction of lines and arrows. For example, Tufte [22] includes the classic example of a flow map by Minard that depicts Napoleon's march to Moscow. Whereas all maps formed a single category, one chloropleth, murder rate (number 22), was grouped with graphs. This is difficult to explain since a similar graph of gas concentration (number 47) was a singleton.

Icons impart a single interpretation or meaning for a picture. Each icon provides a unique label for a visual representation. Icons are used when the meaning of the icon is apparent to the target audience. Icons formed two groups. Multiple-icon highway signs (number 30) and pull-down menu (number 58) formed one group; the IBM corporate logo (number 4) was grouped with photo-realistic pictures. In general, subjects felt the icons in the sample were attractive but that they conveyed very little information (e.g., each icon is a label for one item).

Photo-realistic pictures are realistic images of an object or scene. All photo-realistic pictures and images formed a single group. These representations have a 1:1 correspondence between the real world and the image. Interval properties and distance properties of real world space between objects are preserved in images. Images also have a finite grain

Table 2. Mean Likert scale scores for each of the 11 types of visual representations

Likert scale endpoints	1 spatial 9 nonspatial	1 nontemporal 9 temporal	1 hard 9 easy	1 concrete 9 abstract	1 continuous 9 discrete	1 attractive 9 unattractive	1 parts 9 whole	1 nonnumeric 9 numeric	1 static 9 dynamic	1 a lot 9 a little
structure diagrams	3.0	2.9	6.8	3.3	4.8	4.1	3.8	1.7	3.8	3.7
cartograms	2.9	4.7	5.6	4.7	4.4	4.2	4.7	3.4	4.9	4.3
maps	1.9	2.0	7.5	3.6	4.3	3.9	4.5	3.8	2.4	2.7
graphic tables	6.3	4.7	6.2	4.6	5.3	4.5	3.0	3.7	4.3	2.5
process diagrams	4.7	5.3	5.9	4.9	4.1	4.5	4.1	2.8	6.2	3.9
icons	6.4	2.0	7.3	5.8	7.2	3.9	2.9	2.5	3.4	4.9
time charts	5.9	7.8	6.1	4.9	4.9	3.9	3.7	4.5	4.8	3.7
network charts	5.2	4.0	5.6	5.3	4.3	5.9	3.9	2.6	5.3	4.3
pictures	3.2	1.9	6.7	4.9	5.3	3.1	6.7	1.7	3.1	7.2
tables	7.1	1.8	5.1	5.1	5.4	5.2	2.5	8.0	2.4	2.8
graphs	4.7	4.0	6.3	4.5	4.8	4.6	4.9	7.0	3.7	4.1

size determined by the resolution of the image. As the image becomes too small, objects are more difficult to see and images lose precision. Finally, images have a definite size and shape that limits the amount of the image that can be viewed at one time. Examples include the architectural drawings (number 44) and the robot (number 49). These images were created using state-of-the-art computer-aided design (CAD) software, and other visualization tools. Subjects rated photo-realistic pictures as spatial, attractive, and nonnumeric, and concluded that pictures convey very little information. In fact, subjects felt icons conveyed more information than photo-realistic images.

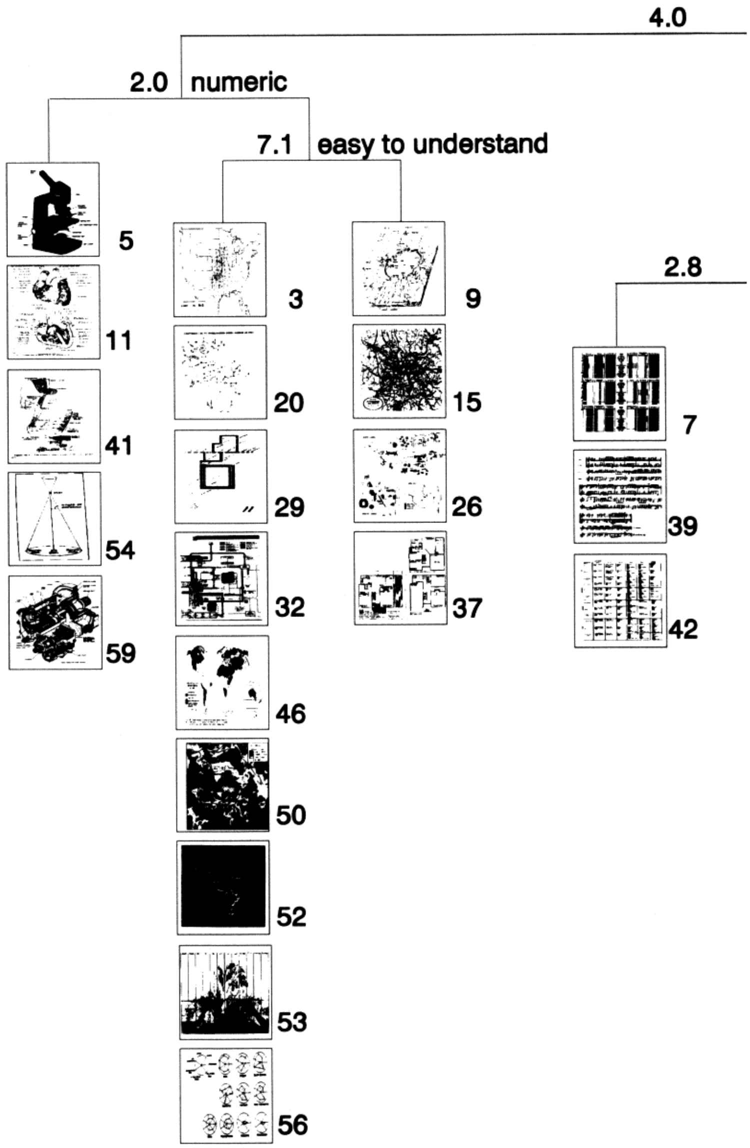
A Closer Look at Apparent Inconsistencies

One of the objectives for developing our classification was to structure this domain of inquiry. By classifying the high-level descriptions of a broad range of visualizations, we can identify inconsistencies. Here we examine these apparent inconsistencies in more detail. In general, we found that many of the apparent inconsistencies have *prima facie* empirical support in the literature, which lends some external validity to our preliminary findings.

Photo-realism often serves as a basis for evaluating the quality of the visual representations. However, subjects in our study characterized photo-realistic images as conveying the least amount of information of all categories in our classification. There are many empirical studies that support this finding. In a series of experiments examining icons, Sorenson and Webb [20] found that recognition errors increased as the photo-realism of icons increased. To enhance identification and memorability, Sorenson and Webb advocate using less complex, more schematic icons. Pezdek et al. [17] found that irrelevant elaborate details about pictures are not easily discriminated from simpler versions of the target pictures without the elaborate details. Marks [15] found the amount of elaborate pictorial detail retained by subjects is a function of the perceptual processes used to encode the pictures. Retention of pictorial details is enhanced when study conditions direct attention to the visual details of the pictures. Thus, pictures may contain a great amount of information, but attention must be directed to the visual details of the picture to enable decoding of this information from the picture.

Although photo-realistic images conveyed less information than all categories in our classification, questions regarding how expert/novice differences influence interpretation also need to be addressed. Certainly the MRI brain image (number 45)

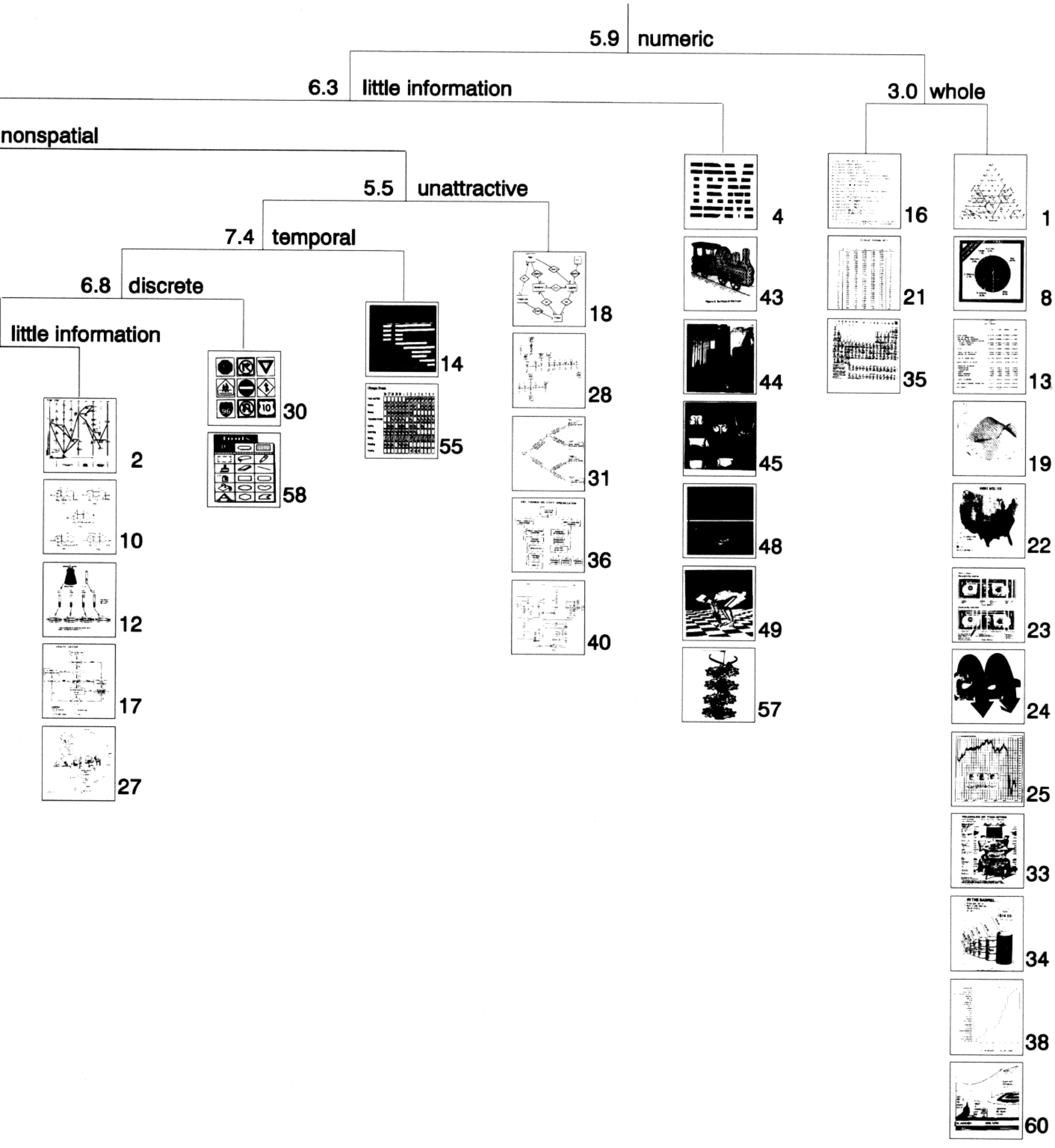
Figure 3. Binary classification tree diagram*



*Graphics go to the left if the cutoff value on the splitting variable is less than or equal to the cutoff value; otherwise, they go to the right. The 10 scale items were (1) spatial-nonspatial, (2) nontemporal-temporal, (3) hard to understand-easy to understand, (4) concrete-abstract, (5) continuous-discrete, (6) attractive-unattractive, (7) emphasizes whole-emphasizes parts, (8) nonnumeric-numeric, (9) static structure-dynamic process, and (10) conveys a lot of information-conveys little information.

would impart more meaning to a brain surgeon than to a college sophomore. The surgeon may have specific schemata for processing information from the MRI display. Novices may not be able to identify the specific areas of the display that contain critical information. These findings suggest that rather than limit a visualization to an exact copy of a real world object, we can enhance photo-realistic images by enhancing the characteristics of some pixels in the image *(smart pixels)* to direct and focus our attention to specific information that is relevant to the current task.

Graphs and tables are often used interchangeably to express the same data. The management information systems literature is rife with empirical studies assessing the efficacy of graphic and tabular information displays [5]. This literature is fraught with conflicting results. Furthermore, recent studies in visual psychophysics discredit some of the sweeping guidelines and generalizations that have previously been reported [21]. Given that we found graphic and tabular representations were more similar to each other than to any other type of visual representation, it would seem more fruitful to examine *how* graphs and tables express knowledge [11, 19] rather than to identify which repre-



sentation is better. We believe this process could be informed by understanding how the symbolic elements of graphs emphasize the whole image and how the alphanumeric entries in tables emphasize parts of the image.

Subjects judged cartograms as being hard to understand relative to either maps or graphs. For example, it was harder to understand the tornado dot map (number 3) than the map of Crater Lake (number 9). This finding has also been demonstrated empirically. Cleveland and McGill [4] found that framed rectangle charts were easier to understand than chloropleth maps because framed rectangle charts facilitate graphical perception. As companies develop geographic information systems for superimposing quantitative and spatial information it is important that designers recognize limitations of cartograms for expressing certain types of information and examine alternative visualization tools for expressing such data.

Two chemical notations, the DNA double helix (number 57) and gibberillins (number 10), were clustered in different categories in our classification. Had subjects considered the use or purpose of these two visual representations, we would have expected them to be in the same group. Subjects judged that the amount of information conveyed was much greater for the traditional chemical notation for gibberillins than for the three-dimensional representation of the DNA double helix. This was surprising given the importance of three-dimensional molecular visualization tools in providing insights for new discoveries in bioengineering. However, it seems more likely that the three-dimensional representations convey more information only to people with an appropriate graph schema for processing information from a novel display format. For example, Winn [24] examined eye movements of perceptual strategies used by subjects viewing normal diagrams and diagrams with unanticipated formats. The absence of an accurate diagram schema for displays with unanticipated formats delayed information processing and caused more information processing errors. Thus, expert-novice differences may not only be a function of graphic arts training but also be a function of having appropriate graph schemata for a particular functional area of expertise.

An earlier study [13] found some differences in categorization of visual representations for subjects with and without graphic arts training. The current study did not find expert-novice differences. Several empirical studies have found expert-novice differences in visual information processing. Wiley [23] found that subjects with graphic arts training remember ordinary pictures better than subjects without graphic arts training, but that memory for unique pictures was consistently high for all subjects regardless of their level of graphic arts training. We might expect to find differences between the memory organizations of graph schemata for experts and novices, as novices often lack the necessary schemata to understand the symbolic notation of the graph. However, DeSanctis and Jarvenpaa [6] have shown that practice and training can improve the ability to decode information from graphs.

Our classification suggests that network charts present nonspatial information that is difficult to understand. It is important to determine how to present spatial information to facilitate understanding. Egan and Schwartz [8] replicated empirical studies comparing memory recall of chess masters and novices using symbolic circuit diagrams. They found that chunking of symbols by experts facilitated their reconstruction of circuit diagrams from memory. As software incorporates spatial displays to facilitate information processing, it is important to understand how chunking facilitates the processing of spatial information.

Temporal data are more difficult to show in static graphics than cyclic data. Given this limitation of static graphics, it may be important for visualization tools to use dynamic displays or animation for analyzing temporal data. For example, smart pixels that highlight key features in computer-generated visual displays could be used to help guide and focus the viewer's attention to particular patterns in dynamic displays.

Summary

Our exploratory research developed a classification from similarity measures for 60 visual representations. There were 11 major clusters of representations: graphs, tables, graphical tables, time charts, networks, structure diagrams, process diagrams, maps, cartograms, icons, and pictures. Our objectives for developing a classification of visual representations were fivefold:

- structure systematic inquiry;
- convey concepts for developing theories;
- identify anomalies;
- predict future research needs;
- communicate knowledge.

By structuring this domain of inquiry at a high level, we can begin to understand how different types of visualizations communicate knowledge. This preliminary structure enabled us to identify some anomalies that suggest limitations of some current visual representations as well as directions for future research.

Our classification is subject to four caveats. First, the sample of visual representations influences how well we can generalize our findings. Had we developed our classification from a larger set of items (600 instead of 60), it is not known whether the 10 Likert scales would still characterize all of the items in the classification. Furthermore, we have not identified deep, hierarchical structure within a cluster. For example, what are the major subdivisions within graphs?

Second, the sample of people whose judgments are used to develop the classification must be representative of the entire range of potential graph users. We have conducted three different experiments using 40 different subjects with a wide range of education, cultural, and graphic arts backgrounds. However, this is still only a small sample from the large population of graph users.

Third, different classification techniques for collecting and analyzing data can and do produce different taxonomies. However, we have used three different techniques over three studies, and each technique has revealed a similar pattern of results.

Finally, our efforts have focused primarily on perceived similarity. We have not investigated whether or not these categories apply to the interpretation of graphics or to the recall of graphical information. For a classification to be useful in both graphical design and research formulation, the classification must represent structure that is used by people in interpreting graphs. This evaluation of our reported classification is our current research goal.

Acknowledgments
We would like to thank Ravishankar Rao and three anonymous reviewers for their helpful comments on earlier drafts of this article.

References

1. Barsalou, L.W. Deriving categories to achieve goals. In *The Psychology of Learning and Motivation: Advances in Research and Theory,* G.H. Bower, Ed. Academic Press, New York, 1991.
2. Bertin, J. *Sémiologie Graphiques.* 2d ed. Gauthier-Villars, Paris, 1967. (English translation: Berg, W.J. *Semiology of Graphics.* Univ. of Wisconsin Press, Madison, 1983.)
3. Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. *Classification and Regression Trees.* Wadsworth and Brooks, Pacific Grove, Calif. 1984.
4. Cleveland, W.S., and McGill, R. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *J. Am. Stat. Assoc. 19* (1984), 531–554.
5. DeSanctis, G. Computer graphics as decision aids: Directions for research. *Decision Sciences 5* (1984), 463–487.
6. DeSanctis, G., and Jarvenpaa, S.L. Graphical presentation of accounting data for financial forecasting: An experimental investigation. *Accounting, Organizations, and Society 14,* 5/6 (1989), 509–525.
7. Dillon, W.R., and Goldstein, M. *Multivariate Analysis: Methods and Applications.* Wiley, New York. 1984.
8. Egan, D.E., and Schwartz, B.J. Chunking in recall of symbolic drawings. *Mem. Cognit.* 7 (1979), 149–158.
9. Jain, A.K., and Dubes, R.C. *Algorithms for Clustering Data.* Prentice-Hall, Englewood Cliffs, N.J. 1988.
10. Johnson, S.C. Hierarchical clustering schemes. *Psychometrika 32* (1967), 241–254.
11. Larkin, J.H., and Simon H.A. Why a diagram is (sometimes) worth ten thousand words. *Cognit. Sci. 11* (1987), 65–99.
12. Lohse, J., Rueter, H., Biolsi, K., and Walker, N. Classifying visual knowledge representations: A foundation for visualization research. In *Visualization '90: Proceedings of the First Conference on Visualization* (San Francisco, Calif., Oct. 23–26), A. Kaufman, Ed. IEEE Computer Society Press, Los Alamitos, Calif., 1990, pp. 131–138.
13. Lohse, G.L., Walker, N., Biolsi, K., and Rueter, H. Classifying graphical information. *Behaviour and Information Technology 10* (1991), 419–436.
14. Macdonald-Ross, M. How numbers are shown. *AV Commun. Rev. 25* (1977a), 359–409.
15. Marks, W. Effects of encoding the perceptual features of pictures on memory. *J. Exp. Psych.: Learning, Memory, and Cognition 17* (1991), 566–577.
16. McCormick, B.H., DeFanti, T.A., and Brown, M.D. Visualization in scientific computing—a synopsis. *IEEE Comput. Appl. Graph. 7,* 4 (1987), 61–70.
17. Pezdek, K., Maki, R., Valencia-Laver, D., Whetstone, T., Stoeckert, J., and Dougherty, T. Picture memory: Recognizing added and deleted details. *J Exp. Psych.: Learning, Memory, and Cognition 14* (3) (1988), 468–476.
18. Rankin, R. A taxonomy of graph types. *Information Design Journal 6,* 2 (1990), 147–159.
19. Rumelhart, D.E., and Norman, D.A. Representations in memory. In *Stevens' Handbook of Experimental Psychology,* R.C. Atkinson et al., Eds. 2d ed. Vol. 2. Wiley, New York, 1988, pp. 511–587.
20. Sorenson, P.F., and Webb, J.M. An experimental evaluation of icon quality. Short talk presented at the *CHI'91 Conference,* New Orleans, La., April 28–May 2, 1991.
21. Spence, I. Visual psychophysics of simple graphical elements. *J. Exp. Psych.: Human Perception and Performance 16* (1990), 683–692.
22. Tufte, E.R. *The Visual Display of Quantitative Information.* Graphics Press, Cheshire, Conn., 1983.
23. Wiley, S.E. The relationship between amount of experience in art, visual perception, and picture memory (transfer of training). Ph.D. dissertation, Ball State Univ., Muncie, Ind., 1984.
24. Winn, W. Perceptual strategies used with flow diagrams having normal and unanticipated formats. *Percept. and Motor Skills 57* (1983), 751–762.

About the Authors:

GERALD L. LOHSE is the Nelson Peltz Term Assistant Professor of Operations and Information Management at the Wharton School of the University of Pennsylvania. His research interests include computational cognitive science, human-computer interaction, visualization, and decision support systems. **Author's Present Address:** The Wharton School of the University of Pennsylvania, 1326 Steinberg Hall-Dietrich Hall, Philadelphia, PA 19104-6366; lohse@opim.wharton.upenn.edu

KEVIN BIOLSI is a Ph.D. candidate in the department of psychology at the University of Michigan. His research interests include decision making and classification. **Author's Present Address:** Department of Psychology, Perry Building, The University of Michigan, 330 Packard Road, Ann Arbor, MI 48104-2994; kbiolsi@mcs.com

NEFF WALKER is an assistant professor of engineering psychology in the School of Psychology at Georgia Institute of Technology. His research interests include graphical perception, interface design, and movement control. **Author's Present Address:** School of Psychology, Georgia Institute of Technology, Atlanta, GA 30332; pswkrnw@prism.gatech.edu

HENRY H. RUETER is a research associate at the Cognitive Science and Machine Intelligence Laboratory at the University of Michigan. His research interests include mathematical psychology and symbolic computing. **Author's Present Address:** Cognitive Science and Machine Intelligence Laboratory, 701 Tappan Street, The University of Michigan, Ann Arbor, MI 48109-1234; henry@crew.umich.edu

This material is based upon work supported by the National Science Foundation under Grant No. IRI-9209576 to first author.