Supporting decision making processes : a graphical and interactive analysis of multivariate data

DOI:
10.6100/IR352075

Document status and date:
Published: 01/01/1991

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

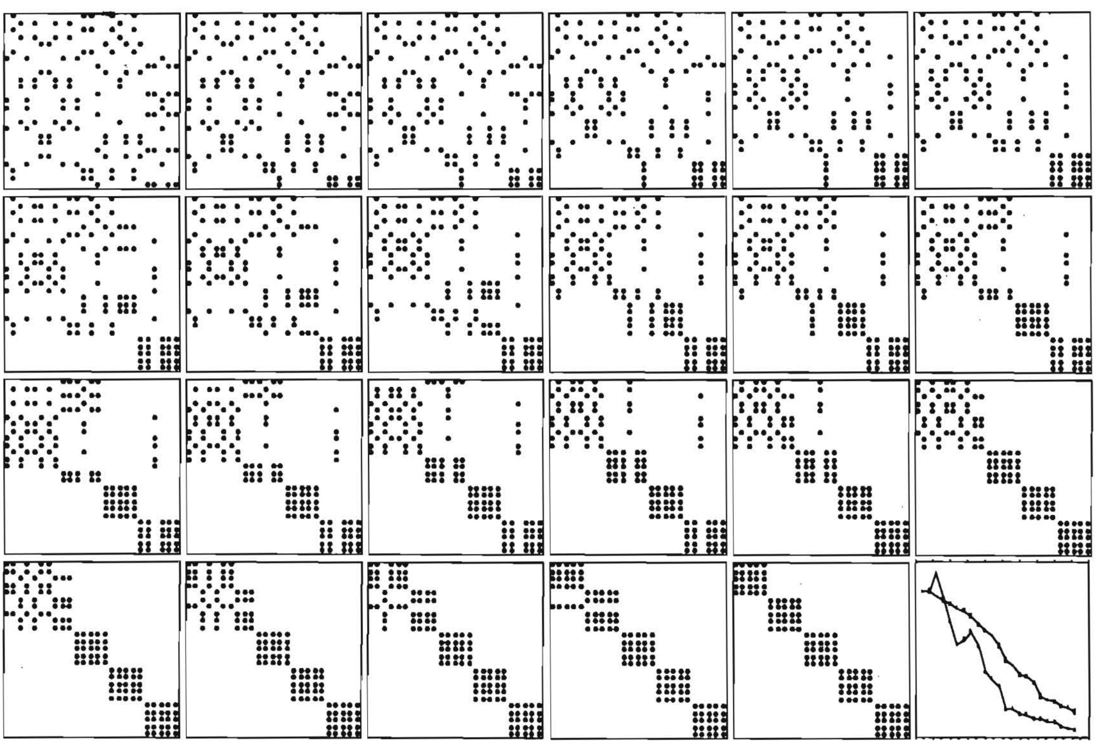
bouwstenen

19

SUPPORTING DECISION MAKING PROCESSES

A GRAPHICAL AND INTERACTIVE ANALYSIS OF MULTIVARIATE DATA

W. ADAMS



faculteit bouwkunde
technische universiteit eindhoven

BOUWSTENEN is een publikatiereeks van de Faculteit Bouwkunde, Technische Universiteit Eindhoven.
Zij presenteert resultaten van onderzoek en andere aktiviteiten op het vakgebied der Bouwkunde, uitgevoerd in het kader van deze Faculteit.

BOUWSTENEN zijn verkrijgbaar bij:

Publikatiewinkel 'Legenda'
Hoofdgebouw 4.92
Faculteit Bouwkunde
Technische Universiteit Eindhoven
Postbus 513
5600 MB Eindhoven

of telefonisch te bestellen: 040 - 472529
040 - 472193

Supporting Decision-Making Processes: A Graphical and Interactive Analysis of Multivariate Data

Proefschrift

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de Rector Magnificus, prof. dr. J. H. van Lint, voor een commissie aangewezen door het College van Dekanen in het openbaar te verdedigen op dinsdag 7 mei 1991 te 16.00 uur

door
Wilhelmus Theodorus Cornelus Franciscus Adams
geboren te Tilburg

Dit proefschrift is goedgekeurd door
de promotoren

prof. dr. ir. J.A.J. Roufs
en
prof. dr. H.J.P. Timmermans

Contents

Chapter 1 Introduction

Although the use of graphics in scientific research boomed enormously since the first examples of this 'lineal arithmetic' by William Playfair appeared in 1786, its use has, for the most part, remained limited to the static presentation of data.

According to Bertin (1981), the possibilities of the graphic method by far surpass the purely static presentation of facts (which he calls graphic communication). He asserts that graphics are also applicable to the interactive analysis and interpretation of information (graphic processing).

However, as an effective tool in the presentation, analysis and interpretation of multivariate and combinatory data, the graphic method has to meet a number of criteria. The purpose of this thesis is to establish and test these requirements. In the present study we are principally interested in the utility of the method in problems of design, planning and decision making in architecture. The required criteria will be investigated by means of experiments on the perceptive and cognitive levels of human information processing.

1.1 The present thesis

As already stated, the theory of Jacques Bertin will play a central role here in. This French cartographer developed a method in which graphically presented multivariate data and their interactive analysis are combined (reorderable matrix). The method was originally meant to support both analysis and decision making with multivariate data and proposes the direct translation of numerical cell values into elementary graphical symbols, such as circles, squares or bars. Analysis and decision making are then based on the visual comparison and interpretation of the graphical elements. As Bertin claims that his reorderable matrix is suitable for presentation and analysis and is effective at different levels of information processing, from individual elements to overall interpretation, the method has to satisfy a large number of requirements touching on these specific purposes. According to Bertin, the fundamental difference between the reorderable matrix method and standard statistical and graphical techniques, such as a t-test or pie chart, is that they have a restricted, rigidly defined purpose, whereas the reorderable matrix could be useful for more divergent purposes. The most important shortcoming of this specific method and, in fact, of large parts of

Bertin's whole theory, is the lack of experimental support for this claim and for many others of his ideas and opinions. The present study investigates the requirements which have to be met if the method of the reorderable matrix is to be a practically useful, and above all, an appropriate tool. This implies, among other things, that it gives a correct reproduction of the *original data* and allows the correct interpretation of differences, similarities and associations. If a number of objects[1] show the same typical characteristics or features, they should be seen as a group of equivalent items, and the rating of an overall configuration (overall picture) of graphic symbol elements in a matrix should correspond to the degree of regularity and association of the data. The ordered appearance of a graphical display gives a certain insight into the underlying structure of the information. The process of *analysing* data with the help of this tool should be straightforward and obvious. *Judgements* and *decisions* based on results obtained with the graphical tool should be verifiable against the original data. The criteria tested in this study touch upon the processing of graphic information from a lower level of perception to a higher cognitive level. After operationalisation they will be tested experimentally for confirmation.

1.2 Information, decision and analysis

If the analysis and the decision-making process are to be based on a number of previously collected and recorded data, at least three important aspects are involved:

- type and amount of available or required information
- required judgement or decision
- type of method, model or technique used in the analysis.

These three aspects are mutually related (as schematized in Figure 1.1).

[1] a correct comprehension of the concepts "object, characteristic and feature" is of crucial importance to the present study. The object concept as it used here, covers both "solid things that can be seen or touched" as well as "things aimed at, intentions or purposes". Thus, all things that have some distinctly visible or measurable features are covered by this concept. Whereas different types of houses, public buildings and factories, but also different persons are all examples of the first explanation, building schemes and town plans, seasons and months of the year are examples of the second meaning. The concepts, "features" and "characteristics" stand for the visible or measurable properties of these objects.

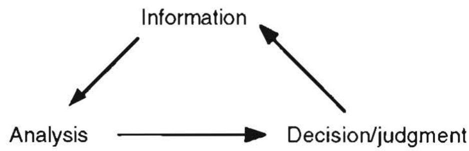


Figure 1.1 Relation between information, analysis and decision

1.2.1 Information

Multivariate versus combinatorial datasets

In his consideration of the available work on problems of architectural design, Höffler (1972) made a broad distinction between two types of information. The first type he referred to in his study, can be designated as data on the dimensions and features of various objects. This type of data can be presented in a rectangular table or matrix and is more commonly known as multivariate data in other disciplines. When for example a number of features is measured or judged for each of a number of alternatives, the results can be recorded in the relevant cells of a rectangular table. For instance, the rent or mortgage, number of rooms, expenditure on electricity and gas, income and age of the householder are some examples of features that can be measured for a number of (alternative) houses and apartments.

Höffler's second type of information can be defined as data on associations and can be presented in a triangular matrix where the cell values represent a kind of strength of association between objects. A common example of data presented in a triangular matrix is the table of inter-city distances in a pocket diary.

In a space planning[2] project, for instance the distances between departments or offices of a business can be reflected in the cells of a triangular matrix in the form of values. By the same procedure, the activity interactions between offices or employees of this business can be recorded (simply by counting the number of times people visit each other in their offices). Combining the data of these two triangular matrices by multiplication of corresponding cells, the total distance that is covered by all employees

[2] space planning is primarily concerned with architectural spatial-allocation problems. These problems can range from the layout of buildings on a large site to the placement of equipment and furniture within a single room. Here, it is especially the arrangement of rooms and facilities within a building that is meant (see e.g. Liggett, 1985; Liggett and Mitchell, 1981; Adams and Daru,1990)

during a certain time can be computed. This value can be calculated for a number of different arrangements of employees and offices. When the criterion is to pursue a minimalisation of the total distance covered by the complete staff, the comparison of these resulting values can serve as the starting-point for the distribution of employees or departments over the building.

Level of measurement

A second possible differentiation of the information is the division into different levels of measurement. In most introductory books on statistics, a differentiation is made between the levels of measurement of nominal, ordinal, interval and ratio data[3]. Information files in certain disciplines in architecture, as in town planning and space planning, often contain a combination of data at different levels (two examples dealing with these disciplines in architecture will be discussed more extensively in Chapter 2). Ordinal data are often strongly represented in these space-planning and town-planning problems, either directly because of the ordinal nature of a feature or indirectly through categorisation of what were originally interval or ratio-level data. In addition to the specific decision to be made, the level of data is a determinant of the type of analysis that is required.

Amount of information

As long as the number of items of information is small, solving the problem should not be too difficult. For example, it will be easy to order the costs when comparing a small number of building projects, by pairwise

[3] In nominal scales each distinct class or group can be assigned a number to act as a distinguishing label, thus taking advantage of the property of identity. As with other scales, all members of a class are regarded as being equal. With nominal scales, the assignment of numbers to classes however is purely arbitrary. Examples of this are types of houses or makes of automobiles. In ordinal measurements, the numbers assigned to the categories utilize the property of rank order. Rank ordering may be regarded as a classification in quantitative categories, where categories are not necessarily equally spaced on a scale. The levels of maintenance of buildings or the utility of the ground for building purposes are examples. The requirement of the interval scale is an equality of units. This means that the same numerical distance is associated with the same empirical distance on some real continuum, for instance the continua of time and temperature. Ratio scales can be distinguished from other scales in that they have an absolute zero point. In ratio scales all fundamental numerical operations are possible and meaningful. Here the examples are numerousness (obtained by counting objects) and size. (See e.g. Guilford, 1954 and Torgerson, 1958).

comparisons. When the quantity of data becomes larger or more complex (the same projects are now judged on 10 or 100 criteria), problems will occur, such as decisions which can no longer be directly deduced from the rough measurements. The data have to be processed more thoroughly to facilitate the decision or selection and results require a clearly structured presentation. Statistical and graphical methods are called for to deal with processing and presentation.

1.2.2 Decision

In addition to the type of information, the decisions or judgements that have to be made are also of major importance when it comes to selecting the right method of analysis. Sometimes investigations are dependent on information already collected. All kinds of data are recorded in databanks nowadays, sometimes by bodies set up for the purpose, such as the Central Bureau of Statistics. The available information could conceivably place some restrictions on the required analysis. If one started with a question or problem and not with already available information, one could actually specify the type and amount of information (and the method of analysis) required in order to answer the question or solve the problem. When the best alternative is to be selected in a multivariate problem, a different type of analysis must be used than when searching for similarities or differences between alternatives. With triangular datasets a travelling-salesman[4] problem requires a different approach (algorithm) than a space-planning problem.

When the starting point is the specific purpose of an investigation, at least three types[5] of problems can be distinguished with multivariate data in a rectangular table.

The first type is that in which the collection and analysis of data should lead to a selection of one or more best alternatives. Examples of this type are the selection of the most appropriate location (site) for a new university (as in example 1 of Chapter 2) or, given a restricted budget, the selection of

---

4 In a travelling-salesman problem, the shortest route must be found where a number of towns, pillar-boxes, businesses, etc. must be visited. End and starting point of the route are one and the same.

5 We do not claim to give an exhaustive list of all types of problems that can be distinguished with rectangular multivariate sets of data, by distinguishing between three types of problems . The three types that are mentioned, however are some typical problems of frequent occurrence within the field of decision making in architectural design.

a number of estates that are to be considered for restoration (one of the objectives of example 2 in Chapter 2). With this type of problem the criteria are considered as of differential importance and are weighted accordingly. The indexed total weight of scores on the criteria determines the relative position of an object in an ordered list.

A second type of problem is that in which the principal purpose of the investigation is classification of the objects. This classification into a small number of existing, predefined typologies has to be based on the specific scores of an object on the examined features. An example of this type is the stock-taking of a number of buildings from different periods (public buildings, factories and houses, all within a restricted area) and the subsequent classification of these buildings into a small number of predefined typologies (also one of the objectives of example 2 in Chapter 2). With this type, normally no differential weights are attached to the criteria, all of which are considered to be of equal importance.

A third area of problems considers investigations of an explorative nature. Here, as with the first and second type, a number of objects is scored on a number of criteria or features. However, objects and/or features are not classified according to some predefined typologies. Instead, the grouping of both objects and criteria is based purely and simply on the similarity and dissimilarity of individual score profiles. Groups in the course of forming do not have to concur with known, predefined typologies, but rather can lead to the discernment of new typologies or to underlying relationships in the data.

Often relationships between features (displayed in the different rows of the table) and objects (displayed in the different columns) can be demonstrated at the same time. A study by Theodorescu (1973) discussed in Bertin (1981) on some 78 characteristics of 82 Ionic capitals, revealed three main types of relationships between groups of characteristics and varieties of capitals. The first type could be characterized by a chronologically evolving change in the design of the capital (becoming more square) and the shaft of the column (becoming wider). The second type could be characterized by the size of the volute (small to begin with, then larger, and smaller again at the end of the period) and an inverse evolution of the overall size of the capital (large at first, then smaller, and finally large again). A third type showed features that seemed unrelated to chronology.

1.2.3 Analysis

As already stated above, the type of analysis is strongly related to the information we have and to the kind of problem we want to solve. The use of statistical techniques in the analysis of multivariate or relational data is often very useful in solving these problems. When the total amount of data is too large to be surveyable, when the data are too complex to yield the required answers directly, or, when in doubt as to whether differences noted really are obvious, statistics can be very helpful. Sometimes a statistical measure is desired simply so that a large mass of data may be compactly summarized. A correctly chosen algorithm or statistical test can indeed reduce a plethora of information into a few important results that, in addition, are appropriate to the specific questions raised or decisions to be made. An algorithm can point out a best choice in a number of alternatives, a chi-square test can fortify the opinion that differences between a number of independent groups are significant or more probably based on chance, and a clustering algorithm can divide a large number of observations into a small number of more or less coherent groups.

When a statistical or mathematical approach is the aim in the analysis of multivariate data, the score on different criteria in the file must be comparable. Consensus must therefore be reached on the relative contribution by the criteria to the final judgement. It is only when consensus can be reached on the weight of all criteria, that the best alternative can be directly estimated. A number of models have been developed for the comparison of data on different levels of measurement. When the problem concerns the selection of the best alternative and the data file contains criteria on several levels of measurement, the usefulness of these models or algorithms could be examined. For a more descriptive analytical approach to the data, nonparametric statistical measures of association, as well as a diversity of clustering algorithms or multidimensional scaling techniques can be used. However, it may be difficult to select the right technique(s), especially when different methods produce results which are at variance with each other.

The analysis of data files containing associative data (triangular matrix) also require their own statistical algorithms. Here the choice of the correct method of analysis may even be more dependent on the type of information available and on the specific decision that is to be made.

1.3 Statistical approach

Statistical and mathematical expedients are becoming more and more common in the analysis of scientific investigations. In the most divergent disciplines (e.g. cultural anthropology, history of art or architecture) investigators take refuge in statistics. As the number of statistical models and tests increases it becomes easier to select one that (more or less) fits the collected data. Yet a statistical approach doesn't always have to be the only or even the best manner in which to analyse information.

1.3.1 Disadvantages of the statistical approach

Sometimes the traditional statistical measures may be unsatisfactory or even ineffective.

- This is particularly the case when it is desirable or necessary to retain a presentation of the complete and detailed set of informations items, for instance, when no consensus can be reached on the relative weight of different criteria or when the decision rules themselves are unclear. Fundamental differences between criteria, for example, comparison of the aspects of cost, aesthetics, safety and environmental pollution show how important, but also how subjective and sensitive the adjudgement of a weight factor to a criterion can be. Moreover, the relative contribution of the adjudged weights can't be retrieved from the final results in most cases.
- When the decision rules are unclear it can, in addition, prove difficult to select between different statistical techniques. When more than one statistical test is possible, with tests showing different results, for example, different approaches in the analysis with clustering algorithms can result in different clusters. Information on the distribution of data and on the number and shape of expected clusters would be helpful for optimal selection of one of the algorithms, but may be unavailable.
- The noncritical use of statistical techniques may lead to overlooking what is perhaps important information. Aggregating scores on different criteria may mean that the strong and weak marks of a specific object add up to an indifferent average. Although this disadvantage can be relevant to the use of statistics, it is not restricted to it but also can be applied to graphic techniques (as we will see in Chapter 3).
- When the researcher has information that is not, or can't be, incorporated in the data file. Bertin calls such information extrinsic.

> "Extrinsic information, that is the nature of the problem and the interplay of the intrinsic information (internal relationships revealed by the image) with everything else. And, by definition, everything else is that which cannot be processed by machine. Extrinsic relationships cannot, by definition, be automated. They are, however, of fundamental importance in interpretation and decision-making. Thus, the most important stages — choice of questions and data, interpretation and decision-making — can never be automated. There is no artificial "intelligence" (Bertin, 1981, p. 9)".

- The investigation could be of a more explorative nature, aiming at the generation of hypotheses instead of testing them (as, for instance in the exploratory study by Theodorescu mentioned earlier). There are no clear-cut hypotheses or questions as a starting point or the investigator has a multiplicity of goals.

When one or more of the above-mentioned arguments applies, it could be desirable to retain a presentation of the complete and detailed set of information items throughout the whole process of decision making, without taking the risk of blotting out information by averaging the observations or pooling two or more characteristics. This detailed display of the information would also facilitate comparison between alternatives and would show the effect of subtle changes in emphasis (the relative weights) of criteria. Subjective preferences (think of more aesthetic aspects) as to the order of the alternatives could easily be compared and discussed.

A tool which would retain the original information intact and as surveyable as possible, while leaving the analysing process to the person in charge, could prove helpful.

1.4 Graphical approach

In addition to a statistical approach in the analysis of multivariate data, a graphical approach is often used for a clear and concise illustration of the ultimate results. Graphical displays are considered suitable for illustrating differences, similarities or associations. It seems only a slight step further to integrate the graphical approach with the analysing process itself. Using graphics during the analysis might be considered instead of using one of the countless graphical formats only when it comes to displaying the results of a statistical analysis. Jacques Bertin actually put this plan into effect and proposed an important role for graphics in the analysis of information.

1.4.1 Requirements of a graphical approach

In order to be actually useful, for example in solving the problems in architecture that were mentioned earlier, a graphical method would have to meet a number of requirements. A broad division into two groups of demands can be made by respectively emphasizing the more communicative (static) and more analytic (interactive) properties of the graphic method. Both demands stress the purpose of supporting the process of decision making. In interactive analysis, all intermediate rearrangements in the layout of the information as well as the final decisions and judgements are strictly reserved to the investigator operating the tool. For this purpose, the total amount of information must be presented in such a way as to display all items visibly and clearly structured.

As regards the first group of communicative requirements, retrieving information on different levels should be feasible. Not only should it be possible to compare individual items but the overall picture should be interpretable at the same time. Another requirement of the layout is a graphical translation or reproduction of the original data that allows a correct interpretation of values, differences, similarities and associations. The graphical translation of the raw data should not distort the original values and their proportions. Both these requirements demand a well-considered but fixed representation of the raw data in a graphical format.

The second group concerns an accessible interactive procedure in the analysis. In order to be really useful in the analysis of data, identical or related values can be grouped in a graphical tool, similar profiles of values of two or more objects can be juxtaposed and compared against a specific criterion, and relations between the row-and-column components can be accentuated by forming groups of elements. To be able to judge the results of a proposed reordering of rows and columns (position of the specific objects in the rows and their features in the columns in the matrix) the matrix configuration really changes and visual feedback on the results of this operation is given. The instructions for desired actions are self-evident and unambiguous.

Bertin has developed a method which, at first glance, seems to satisfy a number of the required options. In the present study, Bertin's method will be operationalised into a number of testable criteria. Results of the tests will be implemented in the method with retrospective effect.

1.5 Contents of the following chapters

Chapter 2 will describe and elaborate some typical problems within the process of decision making in architectural disciplines. The relation between information, analysis and judgement will be illustrated with the help of two practical architectural investigations. The consequences of a possible graphical approach with these examples will be considered.

The historical development of statistical graphics will be illustrated in Chapter 3. Different methods of displaying two-dimensional and multivariate data will be discussed.

Chapter 4 will be devoted to the very elaborate graphical theory of Bertin. Questions regarding the more thoroughly investigated part of his theory (the reorderable matrix) will be operationalised into a number of testable criteria with respect to different levels of information processing.

Criteria at the levels of early visual processing will be discussed in Chapter 5. The required differences in size of matrix elements will be measured, both with stimuli that are presented pairwise and larger groups of graphical elements. In addition to performance of the discrimination and sizing tasks, the ease experienced in performing these tasks will be measured.

The visual estimation of the size of graphical symbols in a matrix will be measured in Chapter 6. Results will be discussed in comparison with fundamental psychophysical research on this topic (e.g. Stevens, 1957). Attention will be paid to the use of a legend or anchor and the effect of this on the estimation of size.

Chapter 7 treats the recognition of patterns in a graphical matrix. First it discusses the validity of a model for the evaluation and prediction of visual clustering on a cartographical map and the utility of this model in a graphical matrix. Then the concept of visual order in a graphical matrix is investigated.

Chapter 8 deals with the actual interactive construction of patterns in a matrix. The process of ordering graphical symbols into more or less coherent patterns is discussed and related to the results of the psychometrically measured concept of visual order of Chapter 7.
In this experiment on pattern construction an automated computer version of the reorderable matrix is used.

Chapter 9 gives a summary and discussion of the present study, together with conclusions reached. The results of the specific experiments and their implications will be considered. Finally, some recommendations for improvement of the reorderable matrix will be proposed.

Chapter 2 Decision making in architectural design

2.1 Introduction

In a number of disciplines or topics in architectural design, such as in town planning or space planning, judgements are repeatedly required which are based on, or at least supported by an often large number of previously collected and registered information items. In processes like research these judgements may take place at different levels

- decisions (e.g. on the geography of buildings and towns or on the feasibility of projects),
- selections (of a best choice between a number of alternatives) and
- findings (when publishing the more general results of an inquiry, confirmations or rejections of hypotheses).

In design processes a similar division can be found in the analysis, synthesis and evaluation phases. In both types of processes the phase or level in the decision making has a strong influence on the type of analysis that can be used.

The required or even possible specific judgement, subsequently depends on the type and availability of information, as well as on the specific test or model used in analysis of such information (as was discussed in Chapter 1).

The relation between type of information, decision and analysis in the process of decision making will be illustrated in two examples. Both are of the multivariate type (discussed in Chapter 1). Similar examples with combinatorial problems could be given. Since the emphasis of the present study is on a graphical approach to the analysis of multivariate data, an example of a combinatorial problem is omitted. However, this does not alter the fact that the graphical approach could not be used with such problems. A number of studies (e.g. Daru, 1989, Adams and Daru, 1990) actually show the opposite.

2.2 Example 1: Selection of a location

In a study by Storbeck (1972) the selection of a suitable geographical site for the establishment of a new university was under discussion. For the benefit of a well-considered selection, a number of possible sites was compared in the light of a large number of criteria. The specific problem in that study was to select the best site by ordering the alternatives. The optimum location would have order number 1, the second best order number 2, etc. Various criteria which were thought to be of importance for the solution of this problem and the decision-making process preceding it were taken into account. A division between micro factors (aspects of the specific areal location) and macro factors (aspects concerning the towns in which these locations were to be found) was made. All criteria on the micro level were measured on a 5-point scale (0-4) as shown in Table 2.1 (note that most of the criteria are ordinal, some are at a ratio level). How the specific values on this micro level came into being is illustrated for two of the criteria in Tables 2.2 and 2.3. The actual assessment of the values of objects (by categorisation) results in specific cell values that are ordinal in nature. When a decision is to be made on this specific problem, the relative weights of the various criteria must be assessed. In that study this was done by allotting the same weight to all measured features. Thus, professional expenses of a site, the distance to major libraries and the distance to other universities were all considered equally important. By adding up the marks obtained on the criteria, the order of the alternatives was easily calculated.

The factors on the so-called macro level were more closely related to the selected towns than to the precise areal location of the university campus itself. Some examples of these factors are the number and average size of houses, the number and diversity of shops, availability and capacity of public health services, recreational and cultural facilities. All selected building locations within the same town were given the same macro value (alternatives 1, 2 and 3 all got the same values as they were all located in Bielefeld). The criteria on the macro level were subsequently treated in the same way as those on the micro level, with all criteria receiving the same weight.

In conclusion, the ultimate choice was determined by adding the two indexed values of both micro and macro levels. It should be noted that modern multicriteria evaluation methods generally use a more complex weighting of the different criteria and are therefore able to give a more carefully balanced appraisal.

In addition to the relatively easy method of analysis that Storbeck uses, some other methods might reveal additional information. With a method that tests the degree of association between features (the y-dimension of the table) the complete set of variables (in this study there were 19 micro-level criteria) might be reduced to a smaller amount without, or with only a slight, loss of information. In subsequent similar research, this might lead to a more precise determination of the information and criteria that are actually required. The more differentiating characteristics could be separated from criteria that show a more uniform distribution of values. The same association tests used on the objects (locations on the x-dimension of the table) could reveal different locations showing similar or opposite profiles of scores. This kind of information could open up discussion on specific choices in the decision-making process. With a clustering algorithm or analysis of correspondence, the complete data set might be grouped into a small number of clusters, with similar values within each of these clusters.

Table 2.1 Values on 19 features for 13 possible sites for the establishment of a university

	sites												
features examined	a	b	c	d	e	f	g	h	i	j	k	l	m
applicable bylaws	1	3	0	1	0	1	0	1	4	4	3	4	4
utility restrictions	2	2	2	0	2	1	2	4	2	3	0	2	2
suitability of soil for farming	0	0	0	1	0	1	0	0	3	1	4	0	3
proprietary rights	0	1	2	3	0	1	1	0	0	3	1	2	3
present-day use of soil	1	3	2	3	0	4	1	2	0	2	4	3	0
accessibility of the site	2	3	0	1	1	0	2	2	4	3	1	2	3
professional expenses	4	4	4	2	1	2	4	2	0	2	3	2	3
accessibility by road	2	4	2	3	3	3	1	3	4	4	3	3	4
served by public transport system	3	4	3	2	2	1	1	3	4	4	3	4	3
degree of use of public transport system	3	4	2	2	0	2	0	2	3	2	3	1	2
importance in regional traffic	3	2	0	3	2	3	0	4	0	4	0	2	2
distance to motorway	3	3	4	0	0	1	3	1	0	1	4	0	1
distance to nearest railway station	1	2	1	2	2	3	2	2	1	1	2	2	3
distance from other universities	1	3	0	2	2	2	0	0	0	2	0	0	4
distance to important libraries	1	2	1	2	1	2	0	2	2	2	0	3	2
urban development plan	2	4	0	3	0	4	1	0	0	2	3	3	4
subject to what hindrances	2	4	0	3	3	1	2	1	1	1	1	1	0
location in attractive setting	2	4	0	4	2	1	0	1	1	1	4	1	3
value of landscape	0	3	1	3	3	3	1	2	3	2	4	1	4
Sum of score:	33	55	24	40	24	36	21	32	32	44	43	36	50
Indexed score:	.60	1.0	.44	.73	.44	.65	.38	.58	.58	.80	.78	.65	.91

legend column names:

a: Bielefeld 1	d: Detmold 1	g: Elverdissen	j: Paderborn 1	m: Paderborn 2
b: Bielefeld 2	e: Detmold 2	h: Herford 1	k: Sennestadt	
c: Bielefeld 3	f: Detmold 3	i: Herford 2	l: Soest	

Table 2.2 Classification of the "suitability of soil for farming" to a 5-point scale

category	value
infertile soil (soil value 15 - 30)	4 points
soil of little value (soil value 30 - 40)	3 points
mixed soil of little to average value (soil value 15 - 50)	2 points
soil of average value (soil value 40 - 50)	1 point
fertile soil (soil value 50 and upwards)	0 points

Table 2.3 Classification of "distance to nearest railway station" to a five-point scale

category	value
easy to walk (up to 1 kilometre)	4 points
within walking distance (1 to 2 kilometre)	3 points
easy to reach by public transport (2 to 4 kilometre)	2 points
within reach of public transport (4 to 8 kilometre)	1 point
distance upward of 8 kilometre; travel time more then 30 min.	0 points

2.3 Example 2: Classification of estates

In a study by Albers (1987), some 63 estates were compared from an architectural and culture-historical point of view. A total of 9 criteria, partly of an architectural nature (e.g. continuity of the function of buildings and park), partly of a culture-historical nature (level of maintenance, amount and quality of historical documentation), were selected for this purpose. The aim of this study was to support policy making in the acquisition, maintenance and restoration of the estates. Thus the purpose of this study was not primarily directed to ordering the alternatives, as in the example of Storbeck but more to a broader evaluation of different methods of analysis as an aid in the process of decision making. The scale of values for the various criteria that were used by Albers differed; for some features a 5-point scale was used (unity of park and buildings), sometimes one or more categories were omitted. For the criteria "uniqueness" and "quality of the estate as an example of the style" only categories 5, 4, 3 and 0 were used, for other criteria only category 1 was left out. In this example, the criteria are again ordinal. Some examples of the categorized values are summarized in Table 2.4. The methods of analysis that were used are described in detail in the study by Albers.

Multicriteria analyses

In the first part of the analysis Albers compared 7 methods of multicriteria analysis. Four of these methods used previously attached order numbers (numeric interpretation method, régime method of Israëls-Keller, rank-order method of Israëls-Keller and régime method of Hinloopen-Nijkamp). The nine criteria (see legend of Table 2.4) were ordered according to their importance. This ordering of the criteria was done by Albers herself. The actual ordering she used was A, B, C, D, E, I, H, G, F. Hence, characteristic A (oldest visible period), was considered most important and characteristic F (continuity of the function of the park) least important. In the other three methods no order number was used (analysis of concordance, numeric interpretation method and régime method of Hinloopen-Nijkamp).

Table 2.4 Some examples of the resulting marks on 9 architectural and culture-historical criteria for 63 different estates

		criterion								
No.	**Name of the estate**	A	B	C	D	E	F	G	H	I
I	Velserbeek	7	4	5	4	4	4	4	4	4
II	Waterland	7	4	5	4	5	5	5	4	4
III	Schoonenberg	4	4	4	4	5	4	5	4	3
IV	Hoogergeest	6	6	3	0	0	4	0	4	3
V	Beeckestein	6	6	5	5	5	4	4	4	5
VI	Huis te Spyk	7	7	4	0	0	0	0	0	5
VII	Kennemergaarde	7	0	0	0	2	4	2	5	2
VIII	Duin en Kruidberg	3	1	5	4	5	4	4	3	4
IX	Burg. Rijkenspark	3	3	5	4	5	4	4	3	3
X	Spaarnberg	3	3	3	3	3	4	2	1	4
.....	....	...	...	...	...	...	...	...	...	...
.....	....	...	...	...	...	...	...	...	...	...
LXI	Huis te Bennebroek	6	4	5	5	3	3	3	2	4
LXII	Reygersbos	6	4	3	4	2	4	2	2	4
LXIII	Swartsenburg	4	4	4	0	4	4	4	3	4

Criteria:
A. oldest visible period
B. most important style period
C. quality of the estate as an example of the style
D. uniqueness
E. unity park and buildings
F. continuity of function of park
G. continuity of function of buildings
H. level of maintenance
I. historical documentation

The effect of not using an order number is equivalent to the method in the Storbeck example where all criteria were given the same weight. Following these two methods (same weights or ordered weights), all multi-criteria analyses resulted in an aggregated score and an ordering of the estates by rank. This means that the raw data and the weights (order) of the criteria were actually used to obtain an ultimate ordered list of the examined estates. Results showed that although the position of specific estates in the ordered lists could vary quite considerably as between the different methods, the greatest changes occurred in the middle part of the ordered lists. In the more extreme parts, all seven methods yielded almost equivalent ordered lists.

Additional analyses

In the second part of the analysis, Albers goes further into the matter of aggregating values to one ultimate ordered list. She states that aggregation is allowed when (1.) a lower score on one criterion can be compensated by a higher score on another, or (2.) the considered scores on different criteria are positively correlated. This means that a high score on one criterion is in general accompanied by higher values on others. Whether criteria are positively correlated was, in the first instance, tested. For this purpose a Kendall's rank-order coefficient of correlation was used. Only some of the coefficients were found to reach values of about 0.50, most of them (56 %) being between 0.0 and 0.1. The results of a subsequently performed multidimensional scaling analysis were also judged to be of little significance. These results led Albers to the conclusion that justification of the aggregation of scores in this example needed to be based solely on the willingness to compensate lower marks on one criterion by higher marks on others.

2.4 Discussion

From comparison of the investigations of Albers and Storbeck we can conclude that the specific questions in both investigations led to a difference in analysis. Storbeck restricted his analysis to a weighting of criteria and measuring an aggregated score because his question was (merely) one of ordering the alternative locations in order to select the best site (the first type of problem that was distinguished in Chapter 1). Albers goes further in the analysis when looking for associations between criteria (second and third type of problem in Chapter 1). For this purpose she used both a rank-order correlation technique and a multidimensional scaling analysis. As these analyses didn't show any obvious results she had to confine her conclusions to the results obtained from the multicriteria analyses. It would be interesting to investigate the reason for not finding any associations. Was this because of the nature of the data or because it would require a different approach to and analysis of the data?

In her discussion of the results Albers notes that some estates produced lower order numbers than she expected (before analysing the results, Albers herself made a list of ten estates that she thought were the most important). She states that although one of these estates had a low level of maintenance, its restoration would be relatively easy. She then continues by saying that in her own preference list she had unwittingly taken into account some aspects that were not included in the criteria, such as 'the possibility of restoration' and 'the value of the landscape'.

Not finding any relations with the help of a statistical approach, not being able to involve extrinsic information in the analysis and missing an opportunity to open up the discussion on specific choices or decisions could be three arguments for trying an interactive graphical approach in these examples.

Chapter 3 Statistical Graphics

3.1 Abstract

The historical development of statistical graphics is followed by a treatment of the various methods of graphically reflecting multivariate data. These methods can be assessed in accordance with the different aspects of the data that they stress, such as the relative importance of the measured criteria for each of the objects or the strength of relations between criteria which apply to the complete set of objects. It is shown that each of the graphical formats has its own advantages and disadvantages as far as the overall analysis of information is concerned and that this characteristic restricts the usefulness of the formats to a specific part of the overall analysis of information. Each graphical technique is optimally suited to answer a specific type of question or hypothesis on the data, but is less useful for alternative analytical approaches. A second general disadvantage of the graphical methods is their primarily static nature. Most graphical formats are only used to communicate the final results of a previously performed statistical analysis of the data and therefore have a more illustrative (communicative) purpose than an analytical one. In contrast with these methods, Bertin's reorderable matrix, which will be discussed in Chapter 4, claims to overcome the disadvantages of both restrictive use and static nature.

3.2 Historical development of statistical graphics

Two major breakthroughs

In the origin of graphical methods, two major conceptual breakthroughs can be identified whose impact remains visible and useful even in present-day graphical presentations[1]. The first was the depiction of mathematical functions in Descartes' coordinate system in 1637. Descartes was the first to systematically combine the principle of coordinates and the idea of mathematical functionality by showing a general manner of characterizing a curve by the relation between each of its points (Funkhouser, 1937). In this way every mathematical equation could be depicted by a curve in the coordinate system. Soon after Descartes' invention, Cartesian curve fitting

[1] A more elaborate account of the historical development of the graphical representation of statistical data can be found in Funkhouser, 1937. Parts of the first two sections of this chapter are based on his historical account.

was used for plotting all kinds of series of physical observations as a function of another series to disclose the laws which governed these related phenomena. In addition to the line chart, the best known and most used graphical format, scatter plots are also based on the Cartesian system.

The second breakthrough originated in England in an economic context as opposed to Descartes' mathematical and physical environment and interests. William Playfair's attention was directed more to simple, but accurate methods of displaying (previously gathered) collections of economic and demographic data. He tried to interpret economic and political trends and to improve the general understanding of these complex matters with a number of excellent and highly inventive graphical formats. In his Commercial and Political Atlas of 1786 and subsequent works, Playfair invented and presented a large part of the elementary statistical graphical repertoire that is still in use today. Of these methods, the bar chart, the histogram and polygon, and the circle diagram or pie chart (see Figure 3.1) are the most important, the best known and most widely used.

A number of other new graphical formats followed in the first half of the nineteenth century. Statisticians and statistical organizations began to realize the unique possibilities of the new graphics for the representation, investigation and analysis of data. These newly discovered graphs and plots include:

- the cumulative frequency curve of Fourier (later called ogive by Galton);
- the normal probability curve. According to Funkhouser this curve was first drawn by de Morgan in 1838;
- different kinds of cartograms, with which the distribution, for instance of various kinds of soil, minerals, religions, types of people were entered on geological maps. Elementary symbols, such as dots, squares and spheres were used on these maps to represent quantities;
- graphical time tables.

No consensus of opinion on standards

At the meetings of the International Statistical Congress (1853-1876) first attempts were made to embody the different graphical and cartographical methods and the signs and colours that they used, in a more uniform theory of graphical representation. Although there was whole-hearted agreement on the need for standardization apart from some minor, vague resolutions, no unanimity could be reached on the recommendations. This interferential dissension persisted at later meetings of the congress and eventually led to the break-up of the International Statistical Congress. For the time being, the layout of maps and diagrams remained in the hands of the experienced designers with their divergent opinions. Even though the International Statistical Congress and its successor, the International Institute of Statistics, were not very effective in drawing up standards governing the layout of maps and diagrams, they had a positive effect on the acceptation and the growing use of statistical graphics. It became more and more common practice to append graphical displays to official publications and it was not long before atlases appeared that were almost entirely devoted to these graphical displays.

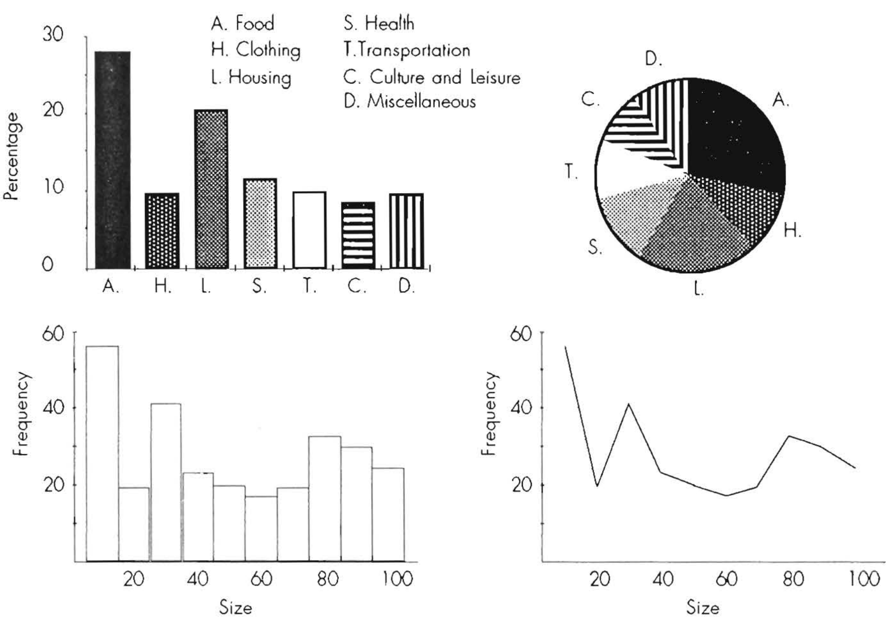


Figure 3.1 Example of a bar chart (top left), a pie chart (top right), a histogram (bottom left) and a frequency polygon (bottom right). In the bar chart and pie chart a fictitious division of a family budget (variable on the x-axis at nominal scale) into seven expense items is given (with expense items in % of total expenditure). In the bottom part, the variable "size" (variable on the x-axis at ordinal scale) is plotted to frequency of occurrence.

Empirical studies testing the merits of graphic presentation

The first studies on the subject of statistical graphics that used an empirical approach appeared in the 1920s. In particular the apparent superiority of bar charts compared to circle and pie charts (comparisons of length were thought to be more easily and surely made by eye than comparisons of area or volume) received a great deal of attention in those experimentally based studies (see Figure 3.1).

In his discussion of the bar-and-circle controversy, MacDonald-Ross (1977) gives an enumeration of these early empirical studies on this subject, with their respective shortcomings and advantages. The general tendency of the results obtained in these early studies is that comparisons based on bar charts are more accurate than those based on circles and squares. These last-mentioned symbols, in turn allow a more accurate estimate of relative size than comparisons based on cubes and spheres. After some decades of diminished research activity on the subject of visual discrimination and sizing, the discussion again flared up in the fifties and sixties (e.g. Stevens, 1957). As aspects of discrimination and sizing of elementary symbols are of importance to the present study, newly accessed insights of this period will be more thoroughly discussed in Chapter 5 on discrimination of graphical symbols and Chapter 6 on visual estimation of size.

Function and declining use of statistical graphs

Fienberg (1979) showed that the use of statistical graphs in two important statistical journals (*Journal of the American Statistical Association* and *Biometrika*) declined almost constantly between 1921 and 1975. The second major conclusion of his study was that the predominant type of utilization of graphics for these two journals was for display and communication, not for analysis. Averaged over the years between 1921 and 1975 and over the two journals, 68% of the number of charts and graphs, and of the space they occupied were used for display and communicative purposes. A total of 15% in number and 13% in space was filled by analytical charts and graphs. The rest of the graphs and charts contained elements of both communication and analysis. According to Fienberg, this decline is largely due to "the relative increase in statistical theory and nongraphical methodology". In this context he notes that pioneering works of Hotelling, Pearson, and in particular, R.A. Fischer were published in the 1920s and 1930s.

Schmid (1983) also mentions the importance of R.A. Fischer's *Statistical Methods for Research* (1925) in the decline of the utilization of graphics. According to Schmid, this book ushered in a new era in which analytical and inductive methodology superseded the more traditional approach. This change in emphasis and orientation inevitably resulted in a downward shift in graphical presentation.

> As the number of statisticians trained in "modern" statistics increased, graphic presentation came to be regarded as more and more irrelevant to what was considered to be the proper domain of statistics. As time went on, this became the predominant point of view and was reflected in textbooks, statistical journals, and courses in statistics. (Schmid, 1983, p. 5)

Although the decline in the use of graphics since 1920 seems obvious, an astonishing increase in innovative graphical formats for the display and analysis of (mainly multivariate) data can be seen, particularly between 1960 and 1980. This innovative boom seems to be closely related to the rise and rapid spread of computers in statistical analysis.

3.3 Graphical representation of multivariate data

The rapid spread of the use of computers in scientific research in the sixties offered the facility of more complex statistical analyses on larger and more complexly designed sets of multivariate data. This, in turn, created a need for statistical graphics capable of dealing with these large numbers of variables and their complex relations. In the case of multivariate data, a large number of features are measured for a large number of persons or objects. One can recall the second example in Chapter 2, where 9 architectural and culture-historical criteria were measured for a total of 63 estates. The most elementary representation of these data is by a data table or matrix. The same purpose can be fulfilled, however, by a number of visually more expressive graphical methods. One way to divide these methods is through the specific aspects of the information they accentuate. Some methods will be discussed in which the objects (glyphs, weather vanes, stars) or features (inside-out plot) are themselves taken as a starting point. Other methods emphasize the relations between these objects (trees, castles) or the relations between their features (symbolic scatter plots and displays). The specific message one wishes to transmit and the particular graphical method selected for this purpose are often causally related. When a specific type of relation has to be demonstrated, a graphical method that is optimally suited for displaying this type of relation is automatically selected. Restrictions of this same method, however, obstruct alternative approaches to the information. An alternative approach accentuating different aspects of the information calls for a different graphical method. In the following paragraphs, some of these methods, with their advantages and shortcomings, will be discussed. In this discussion a restriction is made to two-way displays, both because three-and-more-way displays mostly result in complicated graphs which are difficult to interpret and presenting three or more variables at a time is not a structural solution but, in the case of most multivariate problems, merely eyewash.

Graphics accentuating the examined objects

Some methods have the common characteristic that each of the displayed symbols represents one examined object, with different parts or features of the symbol representing the different features of this object.

A first method of graphically displaying multivariate data is through *glyphs* (Anderson, 1960). A glyph consists of a number of rays emanating from a circle of fixed size with concentric rays (see Figure 3.2). The length of the rays corresponds to the values of the displayed variables, with each ray representing one of them (in Anderson's original version only three different lengths of rays were used). These glyphs are often used as points in a scatter plot. In this way the two most important aspects (and their relation) can be accentuated by depicting them on the x and y-axes of the plot. In addition, the total number of variables that can be displayed is extended to k+2 (with k: number of rays). Such plots, where the two most important characteristics are generally depicted on the x and y-axes, are called *metroglyphs*.

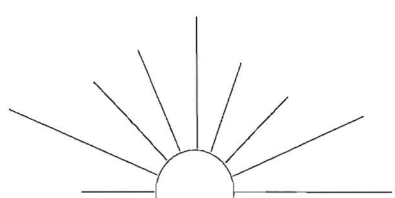


Figure 3.2 Glyph, depicting a number of features of one object. Each of the rays represents one feature, with the length of the rays corresponding to the size of the original values.

A variant on Anderson's glyphs are *weather vanes* (Cleveland and Kleiner, 1974). Meteorologists use these plotting symbols to simultaneously show a number of weather conditions, such as cloud cover, wind direction, wind speed and temperature (see Figure 3.3).

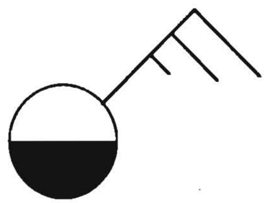


Figure 3.3 Weather-vane symbol. Wind direction and speed are shown by the direction of the flag and the length and number of bars. The shading of the circle shows the amount of cloud cover.

Siegel, Goldwyn and Friedman (1971) also developed a graphical method based on glyphs. They proposed an even distribution of rays over the 360° of a circle as well as connecting the end points of the rays of these glyphs. Originally the rays themselves were omitted, showing only the outline of a polygon in the course of formation. These so-called *k-sided polygons* or *stars* were thought to be more suitable for a quick comparison of the specific distribution of values by comparing the shapes of the resulting polygons (Figure 3.4 and 3.5).

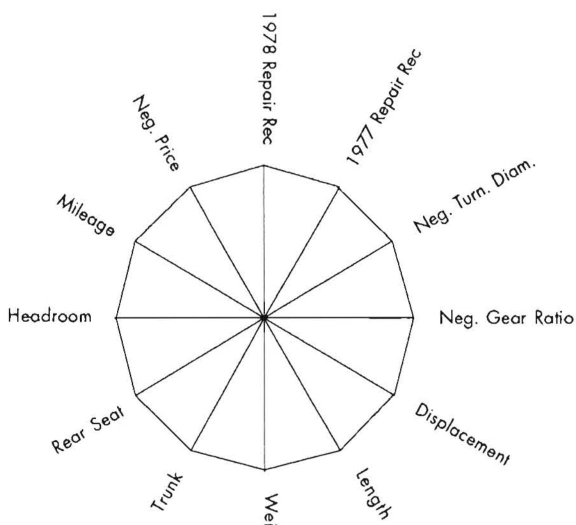


Figure 3.4 Key showing the assignment of automobile variables to the rays of a star (this figure and subsequent figures also using these automobile data are after Chambers et al., 1983).

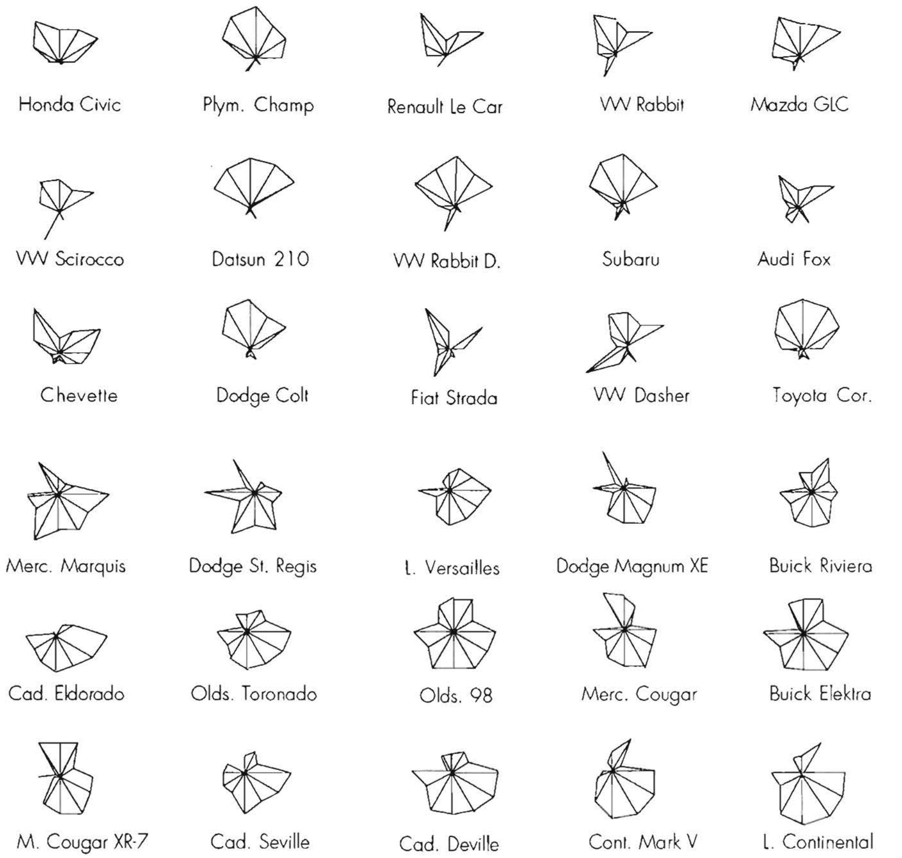


Figure 3.5 Star number plot of all twelve variables of some of the automobile data. Each star or glyph represents one model. Thirty models, 15 light ones (top three rows) and 15 heavy ones (bottom three rows) are shown. This selection is the same as in Figures 3.8 and 3.11.

Andrews (1972) suggested the use of a *Fourier plot.* With this method the k-sided polygon or star is, so to speak, unfolded and the measured variables shown in the stars as end points of the rays (or the principal components instead of the original observations) are here depicted as crests and troughs of a wave in the course of formation (Figure 3.6).

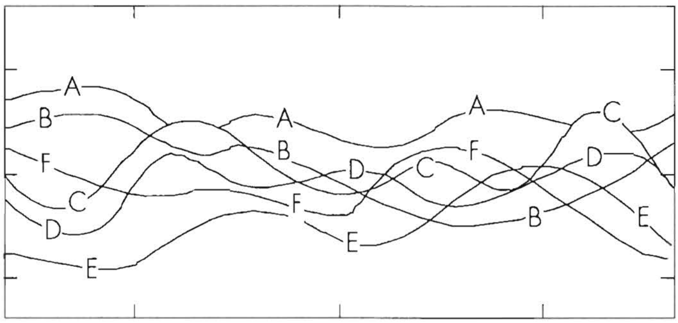


Figure 3.6 Fourier or function plot. Each of the curves represents one object with crests and troughs showing the characteristics or principal components.

In *profiles*, the features are represented by vertical bars, each bar having a height proportional to the value of the corresponding feature. All bars are juxtaposed, rendering a profile of tops and allowing easy and fast comparison of the different heights or the important and insignificant characteristics. This profile format is comparable to a histogram. Another slightly different method connects the bar-top centres. The shape of the resulting format can be compared with a frequency polygon, see Figures 3.7 and 3.8.

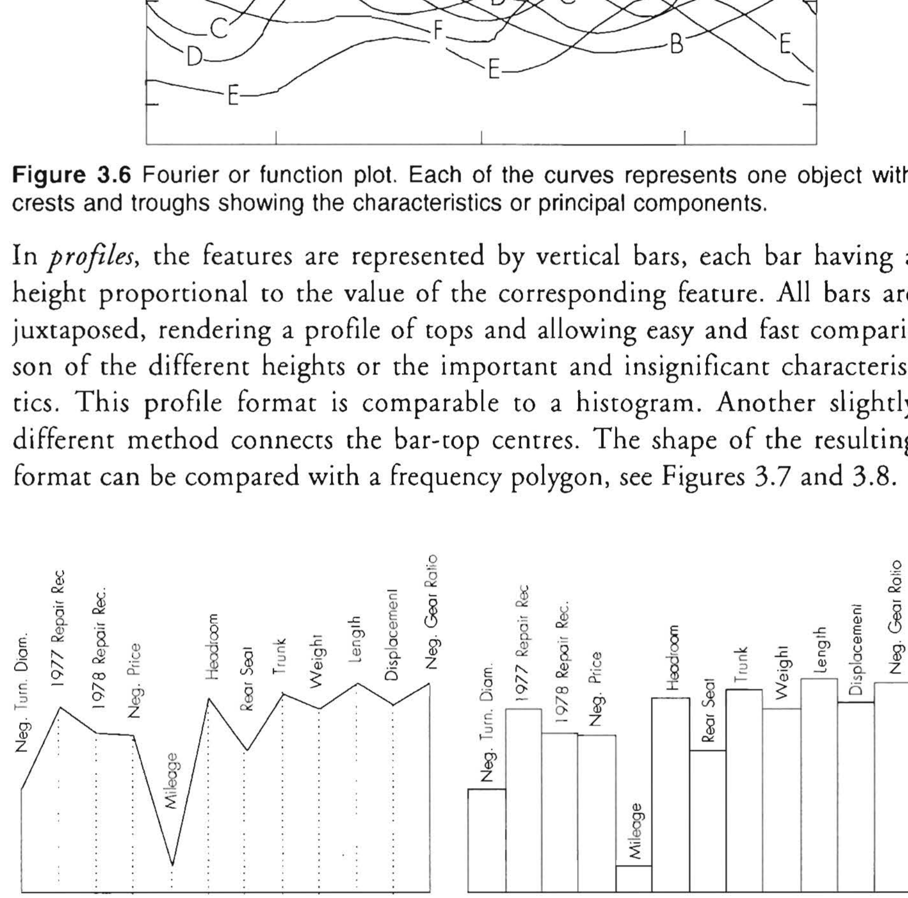


Figure 3.7 Two slightly different versions of the profile symbol. Bars (right-hand part) or end points (left-hand part) symbolise the variables, the height indicating the score on these variables. In this figure the key which assigns automobile variables is shown once more. Here the automobile variables are assigned to different horizontal positions along the profile.

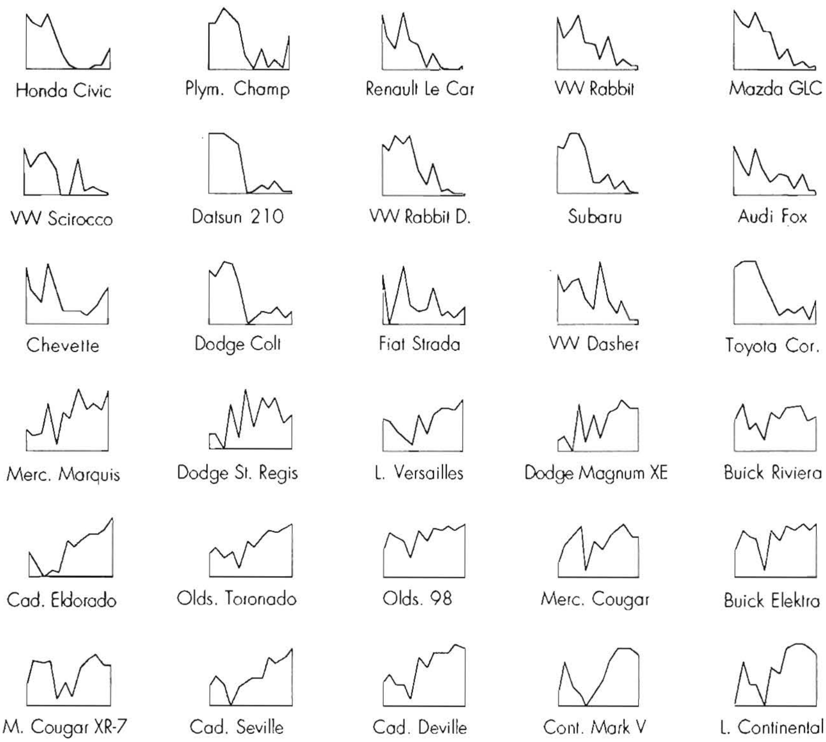


Figure 3.8 Profile-symbol plot of all twelve variables of the automobile data. Each profile depicts a car model. (For different graphical presentations of the same data see also Figures 3.5 and 3.11)

A completely different approach was proposed by Chernoff (1973). He used cartoon *faces* as symbols with data values coded into the facial features (Chambers et al., 1983). In this method, Chernoff relied on human ability to perceive and remember even slight variations in the structure of the human face (Wainer & Thissen, 1981). A quick glance at a face would suffice to serve as a mnemonic device for recalling major conclusions. In measuring speed and accuracy in a card-sorting task, Jacob (1978), found that although cards containing faces or glyphs were both sorted more quickly than those containing digits, sorting faces exclusively was more accurate in addition. According to Jacob, faces not only allowed fast and accurate clustering by visual processing abilities, but also induced their observer to integrate the display into a meaningful whole, a single "Gestalt". An important advantage of this, compared to other methods, is the possibility of detecting slight, barely measurable differences. On the other hand, in circumstances where they are not important, relatively great differences might go unnoticed. Thus noninformative data could be ignored by filtering out insignificant visual phenomena. Useful information could be searched out by focussing on potentially important facial features. Bruckner (1978) considers the subjective assignment of facial features to variables to be an advantage of the Chernoff face method, although he is alive to the danger that its abuse can induce an erroneous impression. As many as 18 characteristics can be presented using features, such as size of the eye and pupil, curvature of eyebrows and mouth, length of the nose and shape of the face.

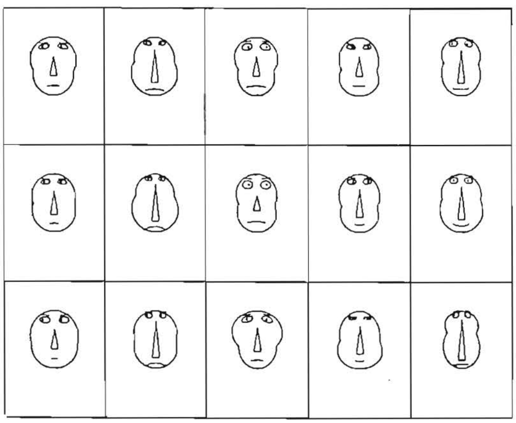


Figure 3.9 Chernoff faces. Each object is represented by a face with facial features showing the different characteristics of the object (after Jacob, 1978).

Graphics accentuating the comparison of objects

In the above-mentioned methods, one distinct symbol was taken as a graphical representation of each object and its characteristics. The position of characteristics in the symbol was more or less arbitrary, although some studies suggest grouping characteristics within the graphics dealing with the same aspects (aspects of cost, environmental aspects, etc.). Systematically changing these positions would, at every turn, result in quite different pictures.

With methods that focus on the comparison of objects, such as *trees* and *castles* (Kleiner and Hartigan, 1981) the position of the features in the plotting symbol is determined by their relation or clustering. In these methods, Kleiner and Hartigan work out a hierarchical clustering algorithm of the variables (they propose a complete linkage or farthest-neighbour method) and use the resulting dendrogram as the basis for their tree-and-castle symbols.

As with the previously mentioned methods, each object is portrayed by one symbol (tree or castle), all trees having the same topology as the tree of variables obtained by hierarchical clustering (see Figures 3.10 and 3.11). Each variable of the data set is assigned to a branch of the tree. The branch length corresponds to the relative value it represents, the thickness is proportional to the number of variables above it in the hierarchy. The extremities of the trees are all of the same width. The angle between the branches reflects the distance between the variables.

The data displayed in Figure 3.11 can be analysed and trees grouped by comparing the contours of complete trees or their main branches. The top three rows of this figure comprise relatively small, thick trees with a marked, large branch protruding to the right. The bottom three rows consist of relatively high, thin trees.

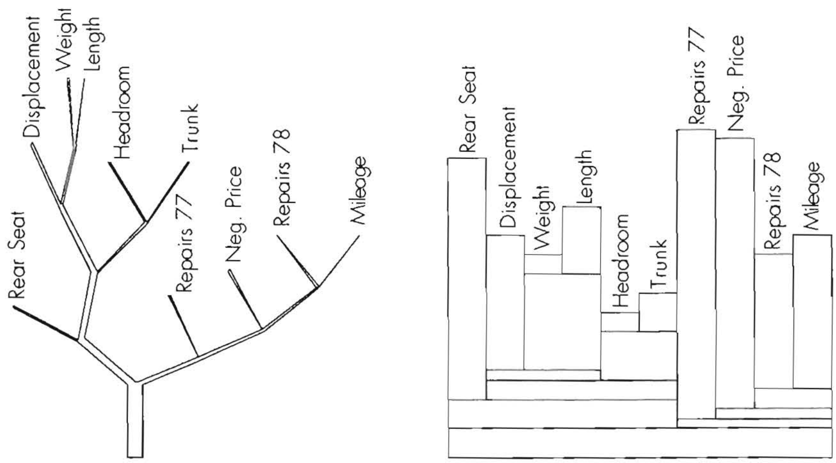


Figure 3.10 Key showing the assignment of ten automobile variables to branches of a tree symbol and to the towers of a castle symbol. The position of the variables within the symbol is determined by hierarchical clustering of the variables.

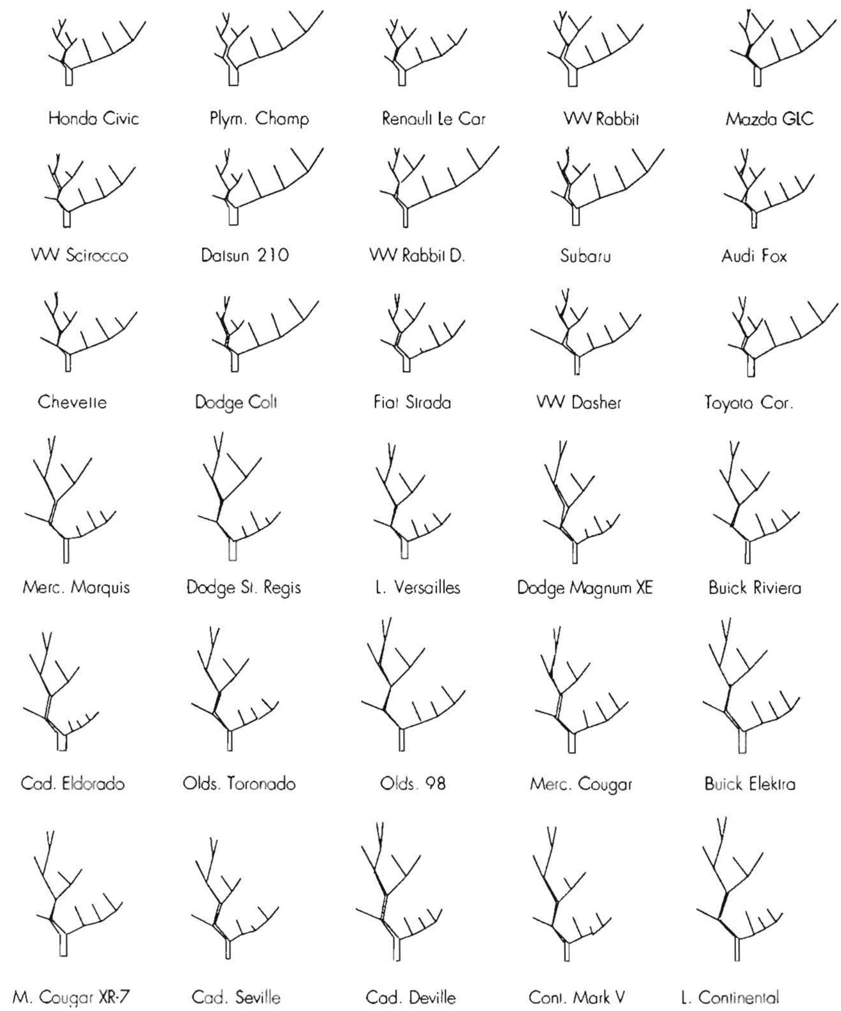


Figure 3.11 Tree-symbol plot of the automobile data. Each tree depicts one car model. For different graphic representations of the same data, see also Figures 3.5 and 3.8.

Graphics accentuating the relation between features

All the above methods have an important characteristic in common. The measured object is considered as a starting point and is represented by a symbol, face or curve. All features of the object are shown by different aspects of these specific symbols, faces or curves. The number of measured objects equals the number of plotting symbols and a specific characteristic is represented by the same feature in all symbols. In addition to these object-based methods, other approaches could also be advocated. Chambers et al. (1983) propose some that are more consistent with the traditional presentation of points or curves in a (Cartesian) coordinate system. Here the different objects no longer form the basis of a more or less abstract plotting symbol. Instead, the pairwise relations between characteristics define the basic unit of the display; the x-y coordinate system. Multidimensional data can be displayed by using different plotting symbols in one (two-dimensional) coordinate system (or window) or by using a multiwindow display. A consequence of this shift from objects to relations between features is that the objects become anonymous in the resulting display. It is no longer possible to directly recover a specific object in a scatter plot unless the plotting symbols are themselves the names of the objects (or there is another one-to-one relation between plotting symbol and object name).

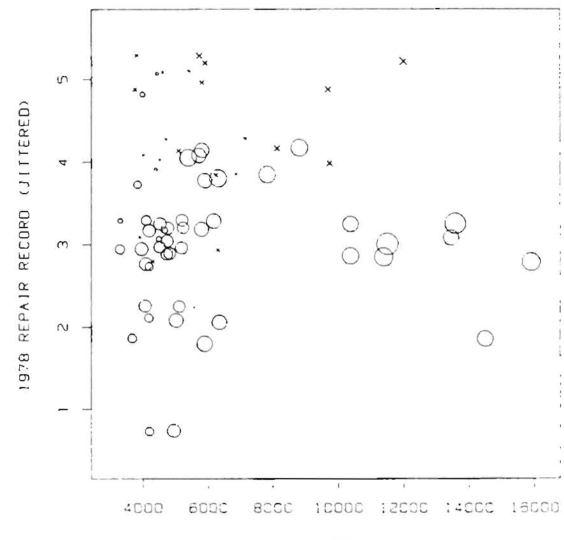


Figure 3.12 Symbolic scatter plot. 1978 Repair record is plotted against price with two variables encoded in the symbols (type of the symbol shows nationality: dots for U.S. cars and crosses for foreign cars, diameter of the symbols show weight of the automobile). After Chambers et al., 1983.

The methods elaborated by Chambers et al. are of two kinds. The first is that of the *symbolic scatter plot* (see Figure 3.12), in which two characteristics are displayed on, respectively the x and y-axes and other features are encoded into the presented symbols, for example by their form, size or colour. A maximum of about 4 to 5 characteristics can be presented simultaneously by this method. In Figure 3.12 can for instance be seen that the larger U.S. cars (bigger dot symbols) are the most expensive ones, as they are mainly located in the right-hand part of the figure, whereas foreign cars (cross symbols) are cheaper (left-hand part of the figure), smaller (small symbols) and score higher on the 1978 repair record (mainly located in the top part of the figure).

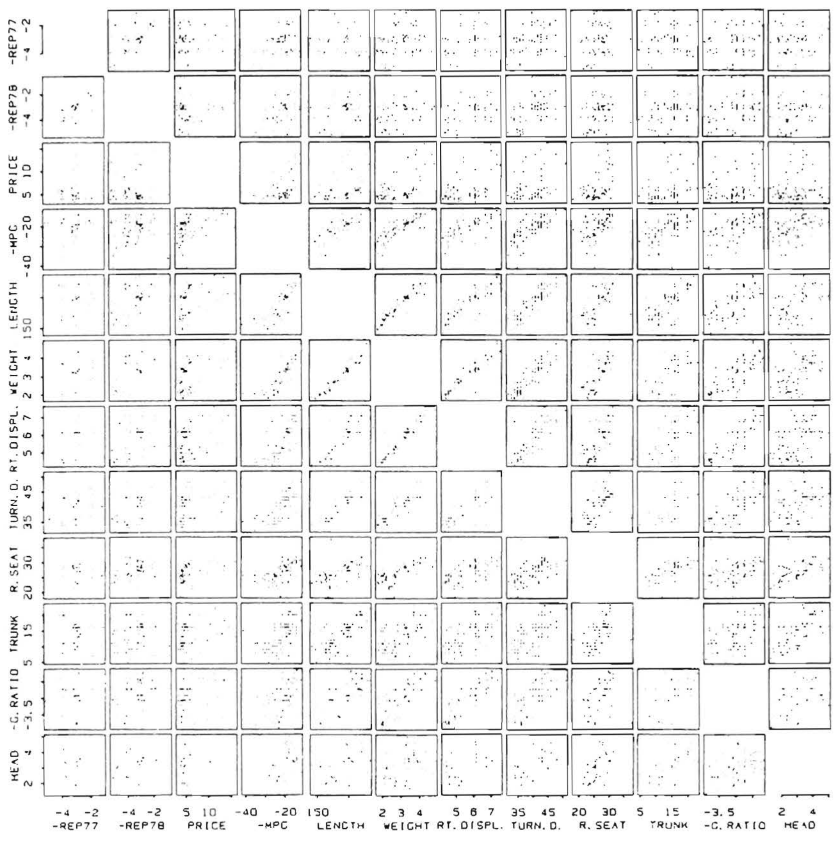


Figure 3.13 Draughtsman's Display. Two variables are shown in each of the component plots. After Chambers et al., 1983.

The second method is that of the *draughtsman's display* and uses multiple windows (see Figure 3.13). Each of the windows presents a scatter plot of a pairwise combination of two features. Thus, a total of (k·(k-1))/2 windows is required to cover all possible combinations (where k equals the number of object features). By combining the symbolic scatter plot and the draughtsman's display, the number of windows needed can be reduced.

The scattered distribution of dots in the first two columns of component plots in Figure 3.13 shows that the variables, repair record 1977 and 1978 are barely correlated with all other variables. The variables, mileage (-MPG), length, weight and displacement (RT. DISPL), on the other hand, all show large mutual positive correlations as the dots in each of matching component plots show a distinct distribution from the lower-left-hand corner towards the upper-right-hand one.

Table 3.1 Inside-out plot showing the scores of ten American states on five "social indicators". In each of the columns the relative positions of the states on these social indicators are revealed by their specific row numbers.

value	Life expectancy	Income	Literacy rate	Population	# Days below 0^0 Celsius
-2.5	PA MS			SD	CA
-2.0			CA	VT	OR
-1.5	AL	MS VT		NH	
-1.0	NH	NH SD	PA MS	IA OR	IO
-0.5	OH				
0.0		IA OR AL	AL OH NH OR		OH PA SD
0.5	SD CA OR	PA	SD IA VT	CA	NH VT
1.0	VT			OH PA	AL
1.5		OH			MS
2.0				MS	
2.5	IA	CA		AL	

AL: Alabama MS: Mississippi PA: Pennsylvania VT: Vermont
CA: California NH: New Hampshire OR: Oregon
IA: Iowa OH: Ohio SD: South Dakota

Graphics accentuating features

A third starting point in the presentation of multivariate data is provided by inside-out plots (Ramsay, 1980). These inside-out plots start from the features. In such a plot, each of the features is presented by a separate column in the table. After the column variables are standardized and row effects are subtracted, the y-component (or row) of the table is used as a scaled measuring staff for plotting the objects. The score of an object on a specific characteristic is marked by entering the name of the object in the specific column and row of the table. All object names can thus be read in each of the columns, the y-position reflecting the original value of the object. This type of plot is especially suitable for the analysis of residuals (Tukey, 1977) and recognition of local extremes (disproportionately high or low values on a criterion). As the absolute position of the object in a column depends on a complex of original values and corrections, the method is unsuitable for judgements on original values or relationships between characteristics.

3.4 Some remarks on two-way displays of multivariate data

A number of aspects of the various graphical methods will be discussed below. The first concerns the number of objects or features that can be displayed by the diverse methods. The second aspect deals with the level of communication and analysis of each method. The level ranges from a retrieval of original individual values to a correct overall impression of the arrangement of all objects and features and their mutual relations. A third aspect involves the different approximations of the graphical methods. Since each of the methods accentuates a specific aspect of the original data, some relations or features are more pertinent than other relations or features. A fourth aspect stresses the specific layout of the plotting symbol and the location of features within it. The fifth aspect touches on the amount of statistical analysis related to each of the methods. Whereas some methods plot the raw data almost directly into a graphical format, others require an extensive statistical analysis of the original information before turning to the actual graphical presentation. The sixth aspect concerns the relation between graphical method and measurement level of the original data. Ability to recognize the specific characteristics and objects in the graphical display is the seventh aspect. The last concerns the empirical basis of the various methods. Only some of the methods appear to have been experimentally investigated.

The number of objects or features that can be displayed

The number of features that can be surveyably displayed depends on the method used. With weather vanes and symbolic scatter plots quite a small number of variables is shown usually (up to approximately 5). Glyphs, stars and Fourier plots can probably handle a maximum of two to three times as many variables, although the strength of the image diminishes rapidly when the number of variables reaches this upper limit. A fixed total of 18 characteristics can be drawn with the original Chernoff faces. Because of the symmetry of the Chernoff face, Flury and Riedwyl (1981) proposed a face in which the left and right-hand parameters could be varied separately. Thirty-six instead of 18 variables can be represented in this new, asymmetrical face. Kleiner and Hartigan (1981) note that up to some 50 variables can be handled by their tree-and-castle drawings.

With all two-way displays that are suitable for the presentation of a variable number of features, there is obviously a positive relation between the required size of the plotting symbol and the number of variables that can be conveniently displayed. With the draughtsman's display, for example, the number of windows to be drawn increases by k-1 with every new characteristic added (k: number of characteristics). The number of columns in an inside-out plot and bars in a profile equals the number of variables of the object. Although the total number of characteristics that can be displayed by these methods is theoretically unlimited, their use should be restricted to the display of a relatively small number of variables, for reasons of surveyability and legibility.

The same comment applies to the number of objects that it is desired to display, as most of the discussed methods use a separate symbol (face, tree, star or curve) for each of the objects. The symbolic scatter plot and the draughtsman's display are an exception to this rule. When there is a relation between two pairwise presented variables, the image becomes clearer, the greater the number of objects that is presented. A weak relation or the lack of a relation between two variables also becomes clearer as the number of objects increases. The same advantage also applies, at least to a certain extent, to inside-out plots. Recognition of individual objects, on the other hand, decreases as the number of objects increases.

Level of communication and analysis

The strength of a graphical format is not restricted to a merely inventive and surveyable graphical display of features and objects, but also offers the possibility of analysing the presented information at different levels ranging from the recognition or estimation of individual values to an overall impression of all values and relations between them.

If they are to be suitable as tools for the analysis of data these graphical methods of representing multivariate data should be capable of answering a number of questions an analyst might ask or show interest in.

- It should be possible to retrieve original values from the chart or graph, or at least have some indication of the relative height of the values. This requirement implies a graphic presentation of all distinct values that were measured.
- Comparison of a specific characteristic shared by two or more objects should be possible, at least at a level of ranking the objects by this criterion, but preferably at a higher level allowing an interpretation of the relative differences.
- The same type of comparison should be possible within objects, that is between features of the same object allowing enumeration of its strong and weak features.
- Comparison of complete objects (that is, of all values of one object compared with all values of another) might be desired. This type of comparison might lead to the conclusion that a specific object has a (noticeably) higher overall score than another object.
- Comparison of a number of features at a time might be desired. Comparing different variables, all touching on a specific aspect (e.g. the cost aspect), might allow quick division of all objects into two or more groups of alternatives (the cheaper and dearer objects). This type of comparison stipulates the possibility of changing the position of the variables within the object. Concentrating the price-related variables in the first quadrant of a glyph or in the left part of a profile facilitates this type of comparison.
- The last two possibilities mentioned could be combined into a simultaneous comparison of all objects and features. By this we mean a division of the total number of objects into a small number of groups, each with a specific division into relatively strong and weak features. This comparison corresponds to an impression of the overall distribution of data, and the ability to divide the total number of values into more or less coherent groups.
- The chart or graph should also be useful in the detection of relations between characteristics (the primary function of trees and castles as well as of symbolic scatter plots and draughtsman's displays). For some graphic methods, however, this requires the option to change the configuration of the characteristics within the graph. When values are standardized within characteristics, comparisons between characteristics are probably directed more to recognition of a possible relation between them.

What pertinent questions can be answered by the different methods?

When different, easily recognizable features of one distinct symbol represent the variables or characteristics of an object, comparison of variables between these objects should not be too difficult. For example, this means comparing the shape of the mouth as between different Chernoff faces or the 270° peaks in a number of stars. Comparisons become more difficult when the number of characteristics increases or objects cannot be properly arranged. When characteristics are organized within a symbol in accordance with a previously performed statistical method, a number of characteristics between objects or even the complete set of objects can be compared at the same time. All peaks in the first quadrant of two glyphs could be compared or a certain large branch could be matched with a number of smaller twigs as between two trees. The possibility of comparing or categorizing complete objects depends on the organization of variables within the plotting symbol and the specific method used. Mezzich and Worthington (1978) in fact, found a number of these graphical representations of multivariate data to be more or less successful when 44 cards, representing archetypal psychiatric patients, had to be arranged into 4 equal groups. Not only did the subjects' success in retrieving the diagnostic groups vary considerably between the different methods, but great differences were also found in the perceived difficulty of the classification task.

Summarizing, these methods are mainly useful for the comparison of specific characteristics or groups of characteristics between objects. In order to fulfil the second possibility of groupwise comparisons or classifications, the desired method has to be well considered and moreover, relating variables should be placed together within the plotting symbol.

With scatter plots and draughtsman's displays it is neither possible to compare different features within the same object nor to match a specific variable between different objects, unless the object names are plotted into the display or are otherwise unequivocally related to the plotted symbols. The main strength of these methods is to provide a general view on the relations existing between two features at a time.

The strength of inside-out plots lies in marking the deviating values. Different variables within the same object can be compared by checking out the vertical positions of the object in the different columns. The same features between different objects can easily be compared within the same column but, once again, these utilities are more appropriate for the more deviating values.

Layout of the plotting symbol

A third observation concerns the location of the features within the plotting symbol. In the case of trees and castles, these locations are fixed by the selected clustering algorithm. This property, however only turns out well if the performed hierarchical clustering technique is indeed the one that best describes the distances between the variables. It might be necessary to compare the results of different algorithms thoroughly before turning to the actual graphical display of the data. As regards Chernoff faces, the assignment of object characteristics to facial parameters needs to be even more carefully considered and should certainly not be capricious. In a study by Chernoff and Rizvi (1975) it was shown that the effect of a random permutation in the assignment of facial parameters affected the error rate in a classification task by a factor of 25 percent. Huff and Black (1978) found the ability of people to replicate a statistically derived typology solely on the basis of perceived similarities among faces to be a function of the assignment of variables to facial features. They claimed that facial differences were perceived on the basis of a small number of special facial features which people deem more important than other features. Size and position of the eyes and the shape of the mouth were considered more important in differentiating the faces from one another than length and width of the nose and size of the ears. When the aspects explaining large parts of the total variance were assigned to features considered to be most important (subjects had to rank the facial features in terms of how important they felt each was in differentiating the faces from one another), the correspondence between actual and expected groupings increased considerably compared to a random assignment of features to the variables. This means that performance in this kind of grouping experiments strongly depends on aspects of the experimental design. In addition, Bruckner (1978) found that, while some of the interpretations of facial features depend only on the input data, other facial features are to some extent interrelated. The results of these dependencies can be deceptive and give an erroneous impression.

The same observation on the location of features within the plotting symbol also holds for the other methods discussed. The expressly chosen and, for the most part, fixed configuration of variables may counteract the message the data structure is trying to convey (Wainer and Thissen, 1981). This problem is usually tackled by careful analysis of the data and the relations between them before passing on to the actual display of the data in the graphical format. Which brings us to the next observation.

Amount of prior statistical analysis

Almost all the mentioned methods require a thoroughgoing analysis and classification (or standardization) of the data before turning to the actual graphic presentation.

Where graphics focus on the objects (stars, profiles, faces), it is not only preferable to categorize the original values of the features (remember that in Anderson's glyphs, originally only three different ray lengths were used), but that features also need to be scaled in such a way that comparisons between different ones can be made. In addition, in the case of some of these object-accentuating methods, the distribution of characteristics over the features of the object needs to be well thought out (and tested) before drawing the symbols (especially with faces). Putting related characteristics together, for example will make the image appear stronger and simplify comparisons between characteristics and objects.

Combination of related characteristics to strengthen the graphic image is precisely what is called for in the methods that accentuate the relation between objects. However, this means that, besides categorization and scaling of the characteristics, some clustering algorithm or other is required in addition to define the relative position of the characteristics in the graphic display.

Methods centring on the relation between features demand relatively little analysis before plotting values into the scatter plot or draughtsman's display. No categorization of features is required and scaling is restricted to calibration of the axis of the plots. The inside-out plot (graphic focussing on features), on the other hand, demands very extensive analysis.

The amount of analysis required before plotting the graphics should be considered in relation to the purpose of the display as an analytical tool. When a number of complex methods is required before a certain graphical method highlighting some aspects of the data can be used, other simpler and more direct methods revealing the same aspects deserve preference.

Level of measurement of the original data

Not every method is equally appropriate to one and the same level of measurement and the actual values of the original data. Most of the methods require (or are specially attuned to) data of an interval or ratio level with values equal to or greater than zero. Features containing exclusively negative values can be "turned over" by multiplying them by a negative value. Moreover, it is preferable to rescale the data so that all variables point in the same direction. A symbolic scatter plot can handle data of ordinal or nominal level (by using plotting symbols of a different shape or colour).

Information on the displayed features and objects

Only some of the graphical methods permit recognition of the different objects and features in the display. Recognition of individual elements, for example is possible in a table by looking at the column or row labels. With other methods the objects and/or features become anonymous in the graphic display. When recognition of an object or feature is desired, these methods require a legend or labelling of the features in the graphic itself. This disadvantage applies to all discussed methods (except to some extent to inside-out plots). Quick and easy shifting of attention to different aspects requires a previously learned legend. The possibility of rearranging features within a symbol acts as a disadvantage. With each change in the configuration of the symbol the legend has to be consulted again.

Empirical basis of the design of the various methods

Although the methods of graphically displaying multivariate data that originated in the sixties and seventies have provided a number of inventive graphical formats (as was the case for the methods displaying more simply structured and univariate data mentioned in the first paragraphs of this chapter), again consensus on a standard display and empirical support of the methods, has, for the greater part not been reached. Some empirical support for Chernoff's faces has been found (Wang, 1978) but at the same time some very serious disadvantages of this method are stressed in a number of investigations. Empirical research on other methods remains practically restricted to some comparisons of different methods within a very restricted task context (e.g. Mezzich & Worthington, 1978). A systematic study on these promising methods is needed.

Conclusion

In summarizing the observations it becomes clear that all the discussed graphical methods have their own specific advantages and shortcomings when it comes to displaying specific features of the raw data or the relations between them. Even though the graphical methods are aimed at a better understanding of the underlying structure of the data, as compared to the numerical tables, they are often unsuccessful in their efforts. The images they present frequently remain unclear and an attendant disadvantage is the prejudice to specific features of the image (Chernoff faces).

We could say that the usefulness of the selected method in displaying or communicating a certain message depends on the specific purpose of the designer. However, in order to be useful as a general tool for the analysis of information, the method should meet some additional generic requirements. When more pertinent and divergent questions on different levels of the information (individual elements, groups, overall impression) can be answered by way of a specific graphic display, then that method is more suitable for a general analysis of data, features, objects and their mutual relations.

3.5 Purpose of the graph or chart

> Of course, there were many reasons other than self-conscious defensiveness for the development and use of "high-powered" statistical techniques. Often such techniques were desperately required. This need provided the impetus for the development of rigorous, formal statistical models, but deemphasized the need for descriptive methods: methods whose aims were exploratory rather than confirmatory. (Wainer & Thissen, 1981, p. 192)

Although Wainer and Thissen claim in their 1981 article (p. 193), that they limit their discussion of graphics to the analytical or processing aspects and refrain from discussing its many other possibilities, including the communicative purpose, they are not clear about the specific and unique aspects of these applications. Whether a graph is used as an analytical tool or its purpose is illustrative seems to be more a matter of diverting the emphasis to one of the functions of a specific (multipurpose) graph than of the actual selection of a graphical method. Whereas their one-way displays of one-way data still show the original data or transformed values and thus in a real sense describe them (thereby directly offering the possibility of aid in the analysis and interpretation), this is no longer valid for some of their two-way displays of one-way data, as in the case of box plots. From their examples we gather that the display of all original data or even showing them in a converted form is not considered (by Wainer & Thissen) to be a prerequisite for usefulness of the method as an analytical tool. When the complete set of original data is not displayed, but only some derived aspects are shown instead, as in the case of residuals in an inside-out plot or a function plot, this automatically leads to a restriction of possibilities in the analysis of the information. Information in an analysis can only be retrieved when it has previously been put into a specific graphical format and is identifiably displayed. The meaning of the word "analysis" in the article by Wainer and Thissen is therefore, at least with a number of graphical formats that they discuss, no more than a static communication or illustration of specific results or characteristics. In addition, a number of examples of analytical graphical formats that are discussed by Wainer & Thissen are also described by other authors, but then as communicative graphics (Schmid, 1983; MacDonald-Ross, 1977; Tufte, 1983; Fischer, 1982).

Schmid (1983, p.3) notes that, in actual practice, the functions of analysis and presentation may be so inextricably interrelated that differentiation is difficult, if not impossible. Fienberg (1979) is more explicit on the distinction between communication and analysis. His division of the communication-analysis continuum into three categories is mainly based on the type of chart, graph or plot. He further notes that "it was clear to us that the placement of graphs into categories was a function of our current perspective and our personal biases". Bertin (1981 p.22), however, seems to be on a promising path in making a distinction between graphic processing and graphic communication

> *Graphic processing* involves two imperatives which do not apply to graphic communication:
> - it must transcribe all the data from the table, that is, the *"comprehensive"* data;
> - it must answer all the pertinent questions and allow the two components of the table to be simplified.
> Graphic processing poses problems of dimensions and manipulations.
>
> *Graphic communication* involves transcribing and telling others what you have discovered. Its aim: rapid perception and potentially, memorization of the overall information. *Its imperative: simplicity.* This simplicity of forms authorizes the superimposition of images. Graphic communication poses problems on the level of simplification and selectivity.

With this division he accentuates the need for analytical graphics as being able to answer all pertinent questions an analyst might ask for, or be interested in. We could regard the purpose of the graphic, the complexity of its underlying data structure and the empirical basis of its method as three distinct gradational aspects of each specific graphical method. The purpose of a graphical technique can shift from a purely illustrative one to a mainly analytical one. The number of variables that are depicted in the graphic are an indication of the complexity of its data structure and the empirical basis of the method varies according to the amount and extent of empirical research on which its practical use is based. Playfair's first statistical graphics were restricted to a small number of variables, were of an explorative nature and mainly served an illustrative purpose. The main point of the diverse multivariate graphical methods, on the other hand, is to depict and analyse more complex data structures whereas their empirical basis still varies considerably from one method to another. In the present thesis we will try to explore the possibilities of Bertin's reorderable matrix method as a communicative and analytical tool and extent its empirical support by investigating and testing the guidelines and rules on which it is based.

3.6 Conclusion

Most of the graphical methods discussed in this chapter that are used to display multivariate data, were initially intended to reveal specific features of the measured objects or similarities, differences or relations between them. These methods owe their usefulness to the specific, often well-organized, way in which they display these features or relations. At the same time, however, their use should be restricted to the particular characteristics that they were intended to display. They are therefore less suitable for an explorative, unbiased or multipurpose analysis of multivariate data. A second constraint of the graphical methods in this chapter is their static nature. It is often very difficult, if not impossible, to reorganize the information once it is displayed in graphical format. Both limitations of these methods are averted in the graphical interactive method of Bertin that will be discussed in the next chapter.

Chapter 4 Bertin's theory

4.1 Abstract

In order to understand the idea of the reorderable matrix method in its context, some general concepts in Bertin's theory and their relations are discussed in this chapter. It is shown that each of the three major types of graphic formats (distinguished by Bertin: the permutation matrices, reorderable network and topography, all have their own specific applications. Selection of the correct type depends on the structure and size of the original set of data and the statistical level of measurement of the objects and characteristics involved. The most thoroughly discussed type of permutation matrix, the reorderable matrix, for instance makes the correct choice when the objects and characteristics are of a nominal nature.

Because of its compact graphic reproduction of all distinct original items of information and the ease with which objects and characteristics can be reorganized, the reorderable matrix permits analysis of the information on different levels, from the level of the individual items to that of the overall picture. In addition, the reorderable matrix allows of testing the a priori hypotheses as well as the a posteriori explorative approach.

In order to transcribe relations of resemblance, order and proportion, graphical signs have at their disposal eight types of variation that the eye can perceive. Of these, the variables "value" (greyness) and, in particular, "size" are shown to be the most powerful when quantitative information on a higher level of measurement is involved. The power of the variable "size" warrants further investigation. This research will be worked out in Chapters 5 and 6.

A number of requirements of an effective reorderable matrix is distinguished and further elaborated in the last section.

4.2 Introduction

Some important principles and starting points of the theory of the French cartographer Jacques Bertin will be briefly discussed here. Bertin's comprehensive theory on the graphical display of data and the processing of this information is enunciated in two books; Sémiologie Graphique which appeared in 1967 and La Graphique et le Traitement Graphique de l'Information (1977). Both have been translated into English in 1981 and

1983, respectively. The most pronounced message in Bertin's theory on graphics and graphical information processing is his plea for the image, whose power he assumes to be largely unused. Both his books are full of illustrations and even statistical principles are, where at all possible, explained, or at least somewhat clarified by graphical means.

According to Bertin, the *data table* forms the fundamental basis of all graphical constructions. In his opinion it is the only construction that always enables one to discover groupings, relations or order along the x and y-dimensions, marked by the z-data.

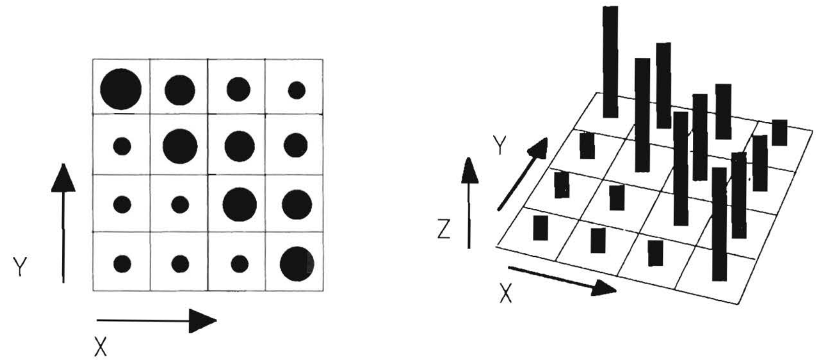


Figure 4.1 The image has three dimensions: x, y and z. Any point in an image can be perceived as the correspondence between a position along x, a position along y and an elevation in z. The set of points can be perceived as the set of correspondences among three dimensions, x, y and z (after Bertin, 1981).

In his theory Bertin therefore uses the data table as a starting point and tries to stick as closely as possible to this specific layout in his graphical approach to data presentation and data analysis. Within the matrix construction, the classification and grouping of similar graphical elements by permutation of rows and columns is the appropriate way to solve information problems graphically. Grouping of elements is based in this case on the highly developed human qualities of pattern recognition and pattern formation. By translating numerical data into visual properties of elementary matrix symbols, on one hand, and allowing a reorganization of objects (the different columns of the table) and their characteristics (the rows of the table) on the other, the original mathematical information problem is converted into a problem of pattern recognition and pattern optimization.

4.3 Permutation matrices, reorderable network, topography

4.3.1 Permutation matrix

There are three major types of graphics around which Bertin's theory of graphics is built and that continually recur in his work, that is the permutation matrix, reorderable network and topography (or ordered network). The first construction and, because of its versatility, the most elaborate and important one, is that of the permutation matrix, which reveals the relationship between two different sets of variables. The elements of these sets are projected as rows and columns on the x and y-dimensions of the table. In the permutation matrix the original data-table layout is preserved by direct conversion of numerical data into graphic symbols.

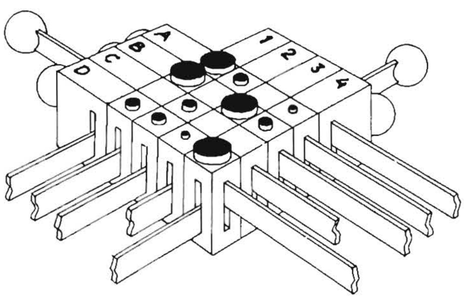


Figure 4.2 Original matrix permutator as developed by Bertin. Rows and columns can be interchanged manually. Intermediate positions were saved by putting the whole matrix on a photocopier.

Reorderable matrix

The most thoroughly discussed type of permutation matrix is the reorderable matrix. In this type, both sets of variables on the x and y-axes are of a nominal nature, allowing a reorganization of the structure of elements by moving the rows and columns of the matrix. In the original version of this dynamic matrix (which was called the "domino" apparatus) Bertin used small plastic cubes that had small symbols painted on top. Rods were threaded through both rows (x) and columns (y) of these plastic cubes. For example, to exchange two columns of cubes, first the x-rods had to be

removed and then the two columns could be switched. Finally the x-rods were replaced.
Intermediate configurations were preserved by placing the complete collection of linked rods on a photocopier (the domino apparatus is illustrated in Figure 4.2).

Because the reorderable matrix maintains the same structure of the data table and also fulfils the requirements of graphical translation and reorganization, it is very suitable for the graphical analysis of information problems. In addition to the reorderable matrix, some other types of rectangular permutation matrices can be distinguished, such as the image file and the weighted matrix. These types also preserve or approximate to the layout of the data table but are more restricted in their reorganization. According to Bertin, the graphical matrix is the only legible way of retaining the overall picture when the number of elements of the sets along the x and y-axes of the table increase. "In data tables of three rows or less, each of the rows can be represented by one of the three dimensions of the image". This type of presentation as in scatter plots directly reveal the relationships and groupings that are present in the data. When the number of characteristics is more than three, the graphical presentation of the data has to return to the fundamental principle of all data constructions, the (graphical) matrix.

Weighted matrix

In the weighted matrix individual values are not directly categorized or translated into a graphical format (Figure 4.3 I). First, the total of each column is indexed to 100% and the percentages of the cells in the columns are calculated (Figure 4.3 II). Next, the percentages obtained are represented by bars and the emanating rows and columns of bar symbols are reordered as in the reorderable matrix (Figure 4.3 III). Now, the width of the bars is made proportional to the original row totals (Figure 4.3 IV).

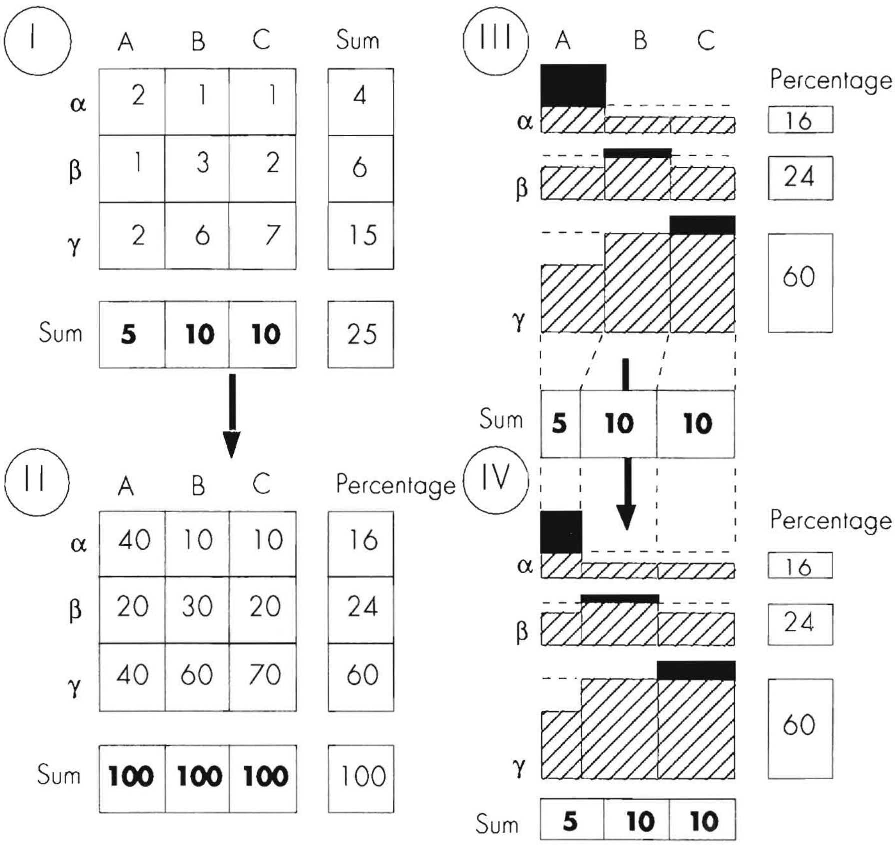


Figure 4.3 Stepwise construction of a weighted matrix. From the original table (I) the vertical percentages are calculated (II). Values are then translated into height of bars and a graphical drawing is constructed (III). Lastly, the width of the columns is made proportional (IV) to the totals per column of the original table. The dotted lines in parts III and IV show the means per row. All black parts of the bars thus represent values that exceed the mean per row (after Bertin, 1981).

As the weighted matrix is a drawn image in which the principle of equally sized table cells is abandoned and width and height of the bars are significant, its use is restricted to relatively small tables. The width of a bar gives an indication of the relative importance of the object, its height represents the relative importance of the characteristic. This construction allows an easy ordering of objects.

Image file and matrix file

The image file and matrix file are constructions for data tables, one of whose components is ordered or contains too many elements to be easily ordered. These types of construction are used to get an impression of the ordering of elements along one of the components, the orderable one. When the number of elements of the reorderable component is not too large (Bertin mentions a number of 30 elements), the categories or values can be represented on a piece of cardboard by small juxtaposed bars (Figure 4.4).

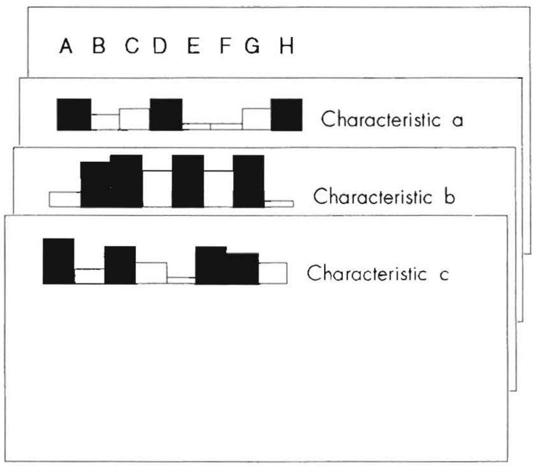


Figure 4.4 Image file. For each of the characteristics (not-ordered component) an object-value profile (ordered component) is drawn on a separate card . Cards can be rearranged according to the images of the different profiles. Black bars represent values higher than a predefined criterion, white bars represent the lower values (after Bertin, 1981).

Bars that are larger than some predefined criterion can be made black (usually the bars representing the values above the mean are made black). These separate cards can subsequently be rearranged according to the specific profiles (images) or distinguishing marks within these profiles. An alternative to the image file is the so-called array of curves. Instead of profiles, curves that connect the different values on the ordered component are drawn on separate cards (the same distinction between profiles and curves was discussed in the preceding chapter and can be compared to the layout of the frequency polygon and the histogram).

With a large number of orderable variables the values of the ordered components can be drawn as strips on the side of a card. With dichotomous data, the presence or absence of a darkened strip indicates respectively the presence or absence of a characteristic; in the case of data of a higher level, the length of the strips represents the height of the measured values. Object names are mounted on separate cards and all cards are lined up (Figure 4.5).

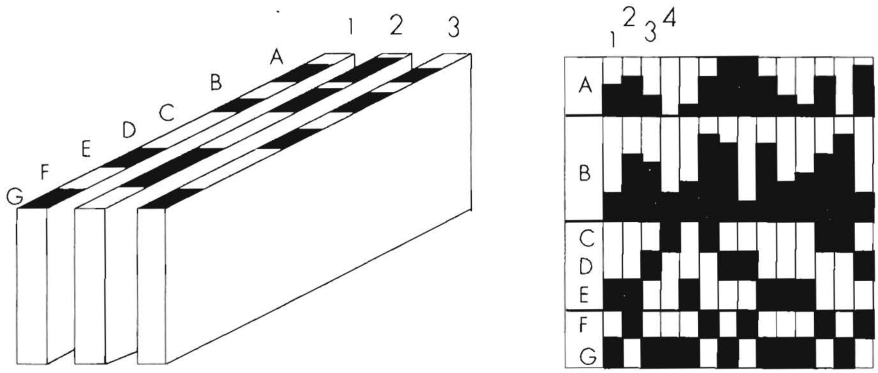


Figure 4.5 Matrix file. For each of the objects (numbers, not-ordered component), values scored on the different characteristics (characters, ordered component) are drawn as strips on the side of a card. The cards are lined up and can be rearranged in the not-ordered (object) dimension. C, D, E and F, G represent groups of related features (after Bertin, 1981).

As with the image file, in this matrix file only the position of a card in the line can be changed. The different strips (bars in the profile) are fixed and identically positioned on each of the cards. In both constructions, pattern recognition and pattern formation can be used in the analysis of the information.

Ordered tables

The last construction belonging to the set of permutation matrices is that of the ordered table. In ordered tables both x and y-components are ordered, so that the position of rows and columns is tied. Normally this type is used when two characteristics, instead of two sets of characteristics are compared. Each of these characteristics, which is of an ordinal or higher level of measurement, is placed along one of the axes of the table. This type of table can only reveal the relationship between two characteristics; a differently organized triangular construction can maximally represent three variables at a time. Other methods have to be found for more than three characteristics, for example the superimposition of tables. This type of construction and its difficulties were mentioned in the preceding chapter where the draughtsman's display and the symbolic scatter plot were discussed.

4.3.2 Reorderable networks

In the matrix representation of a reorderable network, the second construction, there is only one set of variables which is placed on the x and the y-axis of the table. A network shows the presence or strength of associations between the elements. In the paragraph on multivariate versus combinatorial data sets in Chapter 1, some examples of networks or combinatorial data sets were given. Normally only half of the square matrix (the triangular part to the left or to the right of the diagonal) is used to show the presence, or reveal the strength of the relationships among the elements. In order to discover and show the hierarchical structure of the elements, the data table or matrix layout has to be abandoned and other constructions, such as flow charts, organigrams or dendrograms are needed (see Figure 4.6).

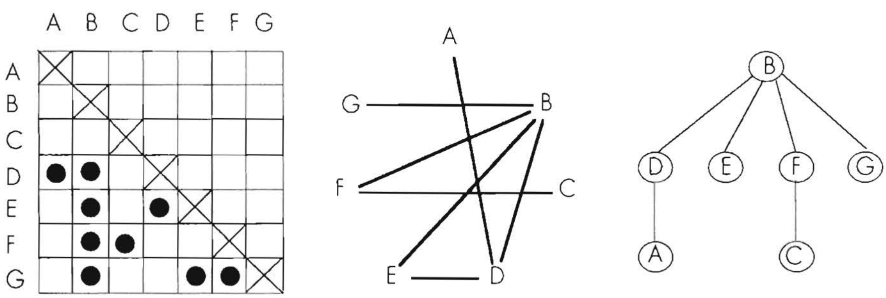


Figure 4.6 Triangular matrix and dendrogram showing the relations and hierarchical structure of some 7 objects (A to G inclusive). The dots in the matrix represent the existence of a relationship between the objects in the relevant row and column of the matrix. In the middle part connections between objects are represented by lines. In the right-hand part the arrangement is visually simplified.

Bertin notes that, although networks are by definition reorderable, just like the reorderable rectangular matrix, their graphical analysis poses more of a problem. The analysis of networks, by graphical representation of the data and subsequent simplification of their elementary structure by transformations of the patterns (comparable with the permutations of rows and columns in the rectangular matrix), proves to be more complex a problem than is the case with rectangular data structures. The actual configuration resulting from permutations of rows is more complex and more difficult to predict in the case of networks than the same type of operation in a reorderable rectangular matrix. In a study on the solubility of graphs, Daru (1989) has, however, shown that the same principles of pattern recognition and pattern formation as used in the reorderable matrix, are also found to function in a triangular matrix. Depending on the type of problem to be solved (e.g. travelling-salesman or space-planning problem) different types of patterns must be sought. In the triangular graphical representation of a network, the problem of the solubility of graphs can be reduced to recognition of two easily identified patterns. When one of these patterns is present in the original configuration of the matrix, or can be created by mathematically correct transformations of the layout, the graph is in principle not soluble, and indirect ways have to be found to remove the junctions in the hierarchical structure.

4.3.3 Ordered networks: Topography

The third type of graphic that Bertin distinguishes is the topography. In topographies, as in reorderable networks, the set of elements can be put into a triangular matrix, with the same set of elements on both axes of the matrix. As opposed to reorderable networks, the elements in a topography are ordered and are not allowed to move along an axis. Cartographical and thematical maps are examples of a topography. In cartographical maps the order is, for instance defined by the geographical location of the individual elements. Solution of the information problem in topographies can therefore only use pattern recognition, as opposed to pattern formation[1], which requires an interactive, reorderable construction of the data. The graphical construction of a topographic network consists of two stages. In the first, topographic stage, the natural arrangement of certain elements of an object are represented on a drawn base map (a set of cities, towns and villages is correctly positioned in a geographical area). In the second, the thematic stage, the theme of the network is added by transcribing the quantity z to each object on the map, with its fixed x and y-positions so that the correct

[1]With pattern formation, we mean the classification of a large number of objects into a relatively small number of clusters or groups. This classification is based on the similarity of the individual objects. This interpretation of pattern formation and pattern recognition is only one of the many intellectual enterprises usually grouped under this heading (Watanabe,1985). Watanabe correctly notes that the classification of objects to unknown classes (as is the case in our interpretation) is an act of cognition rather than a recognition, as opposed to identifying an object as a member of an already known class. "It is taking cognizance of the existence of a group of similar objects" (page 6).

size, colour or shape of each individual city, town and village are determined.

4.4 Stages of graphic perception, types of questions and level of information

Stages of graphic perception

As opposed to the perception of a pictograph, which requires a single stage of perception[2] (what does the presented sign or pictograph signify?), the perception of a graphic requires two stages. The first concerns labelling the information offered in the graphic. What (set of) characteristics of what (set of) objects are displayed by the graphic? What is the topic of the graphic? Bertin calls this the stage of external identification because it requires the identification or isolation of the specific sets of elements out of the inexhaustible domain of possible sets shown in a specific graphic. A drawn figure requires a caption, a table requires row-and-column names, plus a heading explaining the dimensions of the presented image. This external identification must be immediately legible and comprehensible. In addition to this stage of recognition of the displayed sets, graphics demand a second stage of internal identification. At this stage, the true domain of graphics, the relationships among the elements can be discovered. Graphics utilize visual variation and similarity between the signs to indicate relationships, resemblances and differences, or the order of the elements. Visual variables that are suited for this purpose are discussed in a separate section.

Types of questions

According to Bertin there are, in any data table, two types of questions, those pertaining to the x or characteristic component (how many rooms has building A?) and questions introduced by the y or object component (what building has the steepest slope of the roof?). This does not mean that the data table is only suitable for answering predefined questions, taking a fixed level of the x or the y-component as a starting point. As well as testing these

[2]What Bertin calls stages of perception should not be confused with the levels of information processing that are used in psychology. It should be obvious that a number of information-processing levels (such as detection, discrimination, recognition) are required before the meaning of the pictograph is grasped by an observer.

a priori hypotheses, one can also try to detect relationships between objects and characteristics by just looking at the displayed data. This a posteriori or visual method has the potential of revealing quite unexpected relationships or leading to new questions and hypotheses. Analogous to the above-mentioned questions pertaining to the x and y-components we could call this exploration pertaining to the z-component (see Figures 4.1 and 4.7). In addition to making a distinction within the analytical approach between these three different directions in which we can look at the data, another line of approach concerns the level of information one is interested in. The data table can be examined at different levels of the information.

Level of information

At the *elementary* level, one is interested in the value of one single element of the data table. At this level there are as many items of information as there are entries in the table.

At the *intermediate* level, interest is directed to a subset of objects and a corresponding group of characteristics. What are the distinctive characteristics of buildings A, B and C that all belong to the same set of architectural objects? The subset of elements to which attention is directed can be defined in two ways. The buildings in the group or set can all belong to the same architectural style, were all built in the same period or all have the same public function. When interest is specifically directed to the architectural styles, building period or functionality, this means that the subset is specified and verbally defined a priori. The subset can also be constructed by the data. When analysing the image visually, groups of similarly sized elements may strike the eye. In this case the subset is defined a posteriori and visually.

The highest level of information refers to the *overall* image of the data table. This highest level of knowledge that can be attained from the data table concerns the overall relationship between the two components of the data table, that is the set of objects on the x-axis and the set of characteristics along the y-axis. Bertin notes that this overall level is the main purpose of graphics because it is needed for decision making. Type of question and level of information are illustrated in Figure 4.7.

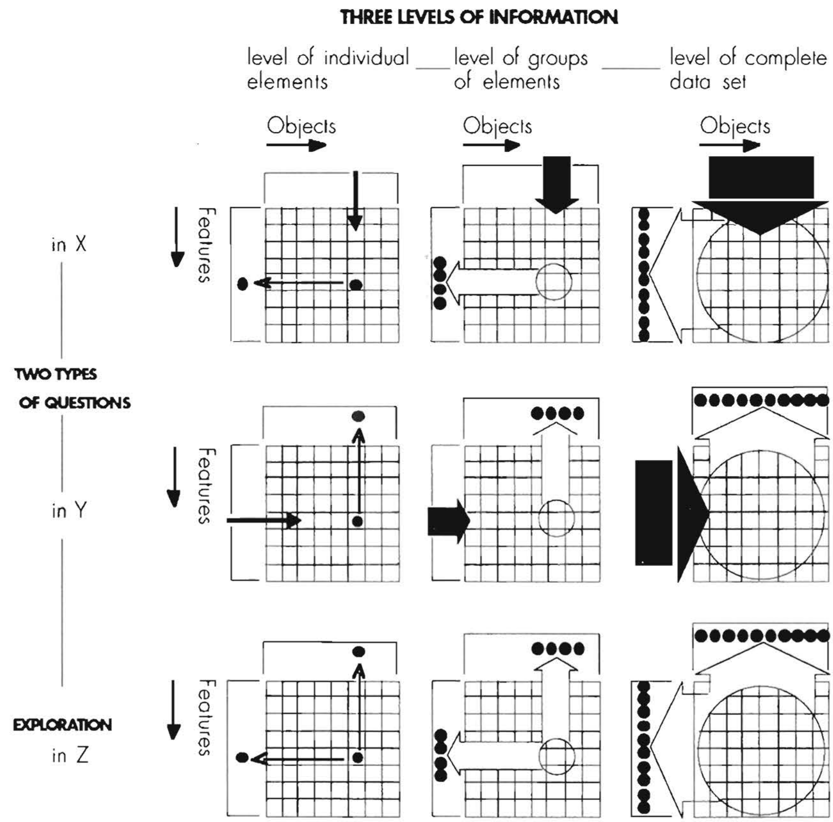


Figure 4.7 Levels of information and types of approach to the information. In addition to the a priori and verbal approach in x and y, the information can also be analysed a posteriori and visually (in z). In the top three matrices, one or more objects are considered as starting-point in the analysis, whereas the analysis starts at one or more features in the middle three matrices. In the three bottom matrices the investigator is guided by the visual image only and abstracts from knowledge of objects and features. The graphical representation of the first two a priori approaches is taken from Bertin. The bottom row of graphical matrices is not given by Bertin.

4.5 Visual variables and their properties

There are some methods in which similarities or differences between signs can be expressed visually. Bertin notes that, in order to transcribe relationships of similarity, order and proportion, graphic signs have at their disposal eight types of variation that the eye can perceive. These eight variables are: the x and y-positions of the sign in the plane, the size and the value of the sign, its texture, colour, orientation and shape. All these characteristics of the displayed signs can be used to transmit information to the observer with regard to the elements presented and their similarities and differences. When two symbols can be visually distinguished for instance by their shape or texture, these variables (shape and texture) can be used to reveal actual differences between objects. These eight variables are not all equally suitable for different types or levels of information. The strength of each of them depends on the type of information they have to transmit. The actual choice of one of these variables should therefore depend on the information that is to be shown. Bertin distinguishes four different properties of the visual variables, comparable with but not completely parallel to the four levels of measurement that are generally distinguished in statistics. Figure 4.8 shows the eight visual variables and their properties.

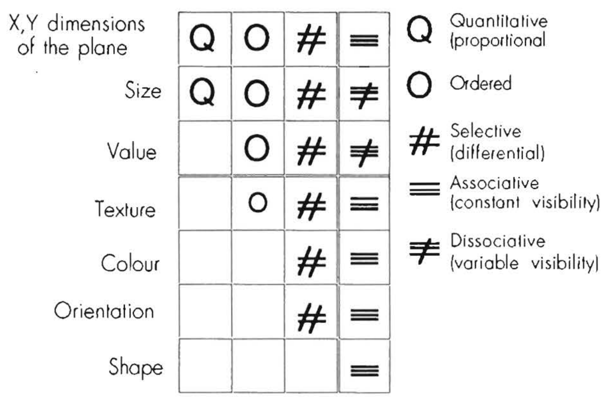


Figure 4.8 Eight different variables of the image and their properties

Association and dissociation

The property lowest in order is that of association and its complementary dissociation. A variable is called associative if all elements can be seen to be the same. For example, in a symbolic scatter plot, where two different symbol shapes are used to present two different classes of objects, an overall indication of the relation between the x and y-components can be denoted when all objects are considered as equal. The total number of elements is treated as one group of equal elements, differences being left out of consideration. With the exception of size and value, all variables are associative. Bertin notes that the reason for being unable to abstract from the properties of value and size visually, is because symbols of unequal size or value also differ in visibility; the larger and darker elements will spontaneously give a visual dominance over the smaller and lighter elements. When the variables of size and value are combined with one of the other visual variables, size and value will again dominate over the others, always resulting in visually different symbols of inconstant visibility.

Selectivity

The second property of the visual variables is that of selectivity. A variable is called selective when all elements of one category (one distinct colour or size) can spontaneously be visually isolated from the other elements. One is able to regard the image that is built up from elements of one single category of interest, by abstracting from elements of different categories. In tests on colour blindness, for example, this property of selectivity (discriminating between objects) is used. According to Bertin, the visual variable, shape, lacks the property of selectivity. When signs of different shapes are mixed within the image, all signs of the same shape cannot be covered in one glance because each of them (target and non-target) is itself a graphical image and therefore needs attention as is illustrated in Figure 4.9. For problems concerning the rapid identification of similar areas (similar in use or quality, not in size) such as are found on thematic maps, shape is unsuitable. In an experiment by Williams (1967) it was found that when subjects knew the shape of the target they had to fix on, the frequency of fixating stimuli of the correct shape was only little above the level expected by chance (see also Engel, 1976). Knowing colour or size of the target resulted in more correct eye fixations.
Besides using shape for the recognition of similarities and differences between elements (nominal level of measurement), the variable of shape, however, has a second unshared efficacy. The property of shape can be used to give an element a symbolic meaning, thus facilitating its external identification. Thanks to this virtue, pictographic symbols are, despite the above-mentioned imperfection, often used on thematic maps. In other applications, such as road signs and pictorial symbols, symbolic meaning of shapes is also used to facilitate external identification. As opposed to shape, all other mentioned properties of a visual element are suitable as differential variables.

Order and quantity

When data of a higher level of measurement (ordinal, interval or ratio level) have to be shown graphically, only five of the eight variables remain available. The variables orientation, colour and shape do not meet the

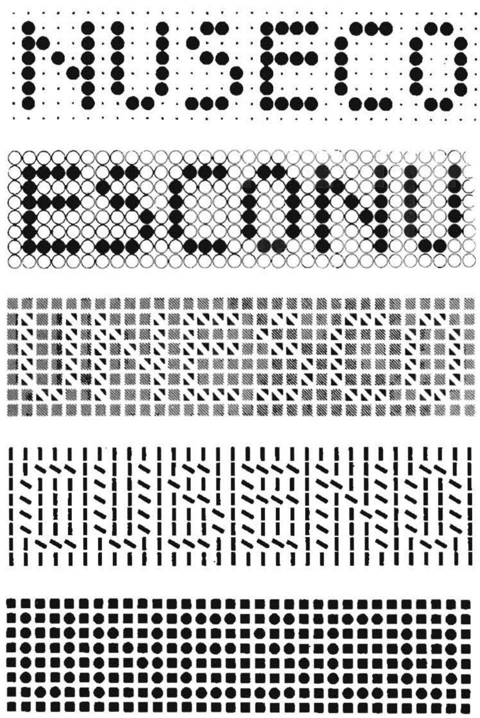


Figure 4.9 Illustration of the associative and selective properties of different visual variables (from top to bottom: size, value, texture, orientation and shape)

requirements of higher levels of measurement and only position in the plane, size, value and texture remain to be utilized to mark an order of categories. In the case of texture, the number of categories to be differentiated depends most of all on the available area. When texture is used to order a number of larger areas on maps, a fair number of categories can be managed; in the smaller cells of a graphical matrix this number decreases dramatically to some three to five, depending on the absolute size of the cells. This makes the property of texture relatively inefficient in displaying data of a higher level. Another legitimate reason for not using texture as a variable feature of elements in the cells of a matrix is because of the illusionary or distorting effects that may appear when a number of areas with different textures are juxtaposed. Since the x and y-positions in the plane are, of course, also impracticable in a matrix, value and especially size seem to be suited to this specific application on a smaller scale.

4.6 Practical application of the reorderable matrix

The first rectangular, adjustable matrix construction is that of the reorderable matrix. Objects in the columns as well as characteristics of these objects in the rows are at a nominal level of measurement.

For example, consider a number of buildings, differing in architectural style, for each of which a number of features are measured. Each of the individual measurements, the contents of the cells of the matrix, can be dichotomous (gable roof or no gable roof), ordered (gable roof with a weak, moderate or steep slope) or numerical (slope of the roof in degrees between 0 and 90). In its most simple form, with dichotomous data, Bertin proposes

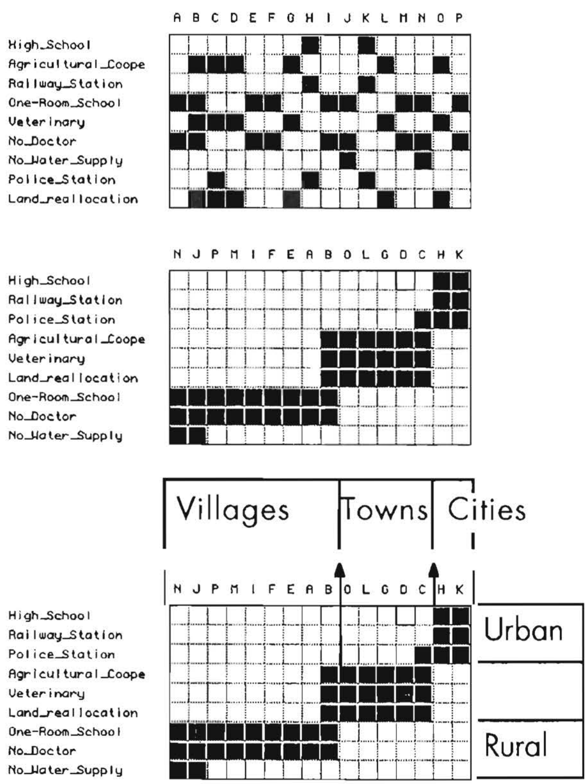


Figure 4.10 Graphical representation and analysis of a dichotomous data table. In the top part, presence or absence of a feature are represented by, respectively, black and white cells. In the middle part, the elements are ordered according to their visual characteristics and in the bottom part the table is clarified by interpretation of the extrinsic information (external identification).

to leave the cells of the graphical matrix empty when the characteristic is absent or make the cell completely black when the characteristic is present (Figure 4.10).

When the data are at an ordinal level or have a numerical value, visual properties of both value and size can be used to clearly indicate the order of the original measurements or represent their numerical values (see paragraph on visual variables). When value is used, the complete area of the matrix cell is "coloured" with a certain greyness, in the case of size, it is an elementary graphical symbol, such as a dot or square, of a certain area which is placed in the cell of the matrix. With ordered data, a higher category should be represented by a larger symbol or a darker cell. Not until the data are at the ratio or interval level, do the precise size of the symbol, the greyness of the cell and differences in size and greyness become important. The visual estimate of relative size or greyness should be equal to the actual numerical (or categorized) values they represent and to the differences or ratios of said values.

According to Bertin, the difference between two (greyness) values is not equal to the simple ratio of two amounts of black but to a ratio of ratios. He claims that the value of a cell that has a grey level of 20% will not be estimated as twice the value of a cell that has a grey level of 10%, but as 2.25 times as much, because this equals the ratio of black to white of the first cell (20/80) divided by the ratio of black to white of the second cell (10/90). When the grey level shows a uniform increase (as in the series 10%, 20%, 30%, 40%, etc.) we will not get equidistant estimated values between the elements of this series. The estimated differences at the upper and lower ends of the scale will be greater than in the middle of the range. Only with a uniform increase in the ratio of ratios will an equidistant sequence be obtained. The estimated quantities corresponding to this corrected sequence show an arithmetic progression (a uniform increase in simple ratios of grey levels will result in geometric progression of estimates). When the lowest value (ratio of black and white or grey level) is known as well as the number of steps required, the consecutive values can be calculated as

$$r = \sqrt[n-1]{\frac{S}{W}} \qquad [4.1]$$

where r: the ratio of ratios of two consecutive steps,
n: the number of steps required
S: the grey level of the highest value
W: the grey level of the lowest value.

In the case of size, the perceived difference between two signs equals the simple ratio of their areas. Bertin notes that "whenever this ratio conforms to the general law of perception, constant proportionality, the surface area of the signs has a geometric progression" (1981, p.198). Estimating the areas of a geometrical range of symbols, a corresponding logarithmic progression can be expected. Bertin notes further that the absolute size of the signs in an accumulation (as, for example on maps or in a matrix) must be "sufficiently" large if they are still to be perceived as distinct signs and not as the shading or grey level of an area.

Within a series of symbols where the surface area of the largest one is ten times that of the smallest, a total of 20 steps provide slight but still perceptible differences. This means that the areal ratio of two successive symbols in a series must be at least $\sqrt[20]{10} \approx 1.122$. The series can be extended at either end with two restrictions toward the bottom. The first restriction is the above-mentioned absolute size of a symbol that is required to be still perceived as a distinct symbol. The second restriction concerns the threshold of differentiation. When the area of a symbol decreases below a certain absolute value (Bertin does not state which value) the areal difference from the next symbol in the series ($\sqrt[20]{10}$) drops below the threshold of differentiation, so that greater differences are needed. How much greater the difference must be is, again, not revealed. The subject of the threshold of differentiation will be further elucidated in the next chapter.

4.7 An experimental approach to the reorderable matrix

Theory and empirical support

In contrast to the theoretically restrictive methods of graphically displaying multivariate data that were discussed in Chapter 3, Bertin's method of the reorderable matrix is based on an extensive theory of graphics. In his theory, Bertin discusses a number of visual variables and their specific properties, different types of graphic construction with their own sphere of activity, diverse levels of information in the analysis of data and differences between external and internal identification. Last, but certainly not least, Bertin is one of the very few who makes a clear-cut distinction between the graphical functions of display and analysis.

Compared to the methods in Chapter 3, on the other hand, Bertin's methods also lack empirical support. There is no experimental verification of many of the rules mentioned. This lack can be explained in part by assuming that Bertin simply doesn't mention the sources he used. From examination of some of the rules, it may be concluded that Bertin really is informed about research that has been done on the specific subject. To some extent, however, sources are missing because no explicit research on the subject has been accomplished (at least, could not be found). These rules are assumed to be based purely on experience. Particularly those theoretical parts can be queried that are concerned with the analysis of the graphical matrix by recognition and construction of visual patterns.

In order to corroborate and test the theory of the reorderable matrix in particular, some important aspects of the graphic presentation and analysis will be examined experimentally.

Man-machine interaction

The "man-machine interaction" approach has been chosen for investigating the reorderable matrix method. On the "man" side of this approach, human possibilities and restrictions in the task context or environment must be taken into account (Falzon, 1984) in which the levels of visual perception, motor operations and cognitive information processing are the main concerns. On the "machine" side, the possibility of automating and computerizing Bertin's original manual domino method (see Figure 4.2) will be investigated by using a computer system with an interactive graphical display. In the task of graphical information processing, it is important to know what part of the task can be taken over by the computer, and what is best left in the hands of the user (Daru and van Gils, 1987). Both "partners" have specific capacities and shortcomings; the latter should be mutually compensated, the former mutually enhanced. In this interaction a need can be distinguished for compatibility between operator and system, which is effective on three levels, that is early visual processing, motor operations and cognitive processing of information. Detection, discrimination and recognition of visual information depend heavily on physical characteristics of the graphical elements in the display. On this basic, perceptive level the compatibility between stimulus characteristics and visual perception should be investigated. On the second level, compatibility between man and machine with regard to operations that people have to, or want to carry out has to be ensured. Exchanging rows and columns of the matrix and other modifications in the configuration of matrix elements should be realised by means of simple, obvious operations. The computer system must be optimally adjusted to such operations. On the level of cognitive compatibility the stages of human information processing (structuring), the strategies that are used in the structuring process and the mental representation of the image must be isomorphic with the structure of information in the design of the computer program.

> Reaching the fuzzily defined goals attached to to these [perceiving, memorizing, decoding and interpreting] tasks is an iterative and dynamic process. People use various strategies to attain these goals. It is essential that computer systems should be as compatible as possible with the actual thinking and problem-solving strategies of real-world users. (Daru and van Gils, 1987, p. 920)

Requirements of the tool

Having distinguished these three levels of compatibility, criteria need to be developed and tested that measure the relative suitability of the method of the reorderable matrix and its actual computerized implementation. At the first level there is, for example, the extent to which the display aids in the sensory processing of the information. At the second level, a possible criterion could be the degree to which people are able to carry out elementary structuring operations, or the rate at which this can be done. A more general criterion for evaluating the relative effects of different display methods could be the extent to which the display provides scenic realism or gives a visually comfortable experience.

In Chapter 5 the proposed size ratio of 1.12 that Bertin considers necessary for the discrimination of elementary graphic symbols will be examined. In the same chapter it will also be necessary to investigate whether this ratio is sufficient for ordering larger sets of graphical symbols. Actual performance on this sorting task will be compared to the complexity experienced in its execution.

Chapter 6 deals with estimation of the size of graphical symbols in a matrix. Although this task is more complex than those in Chapter 5 (because it requires the interpretation of size and differences in size between objects) it is also a task that is mainly enacted at the level of visual perception. As mentioned earlier, the graphical translation of the original or categorized data should allow correct interpretation of values and differences. The graphical translation should not distort the original values and their proportions.

Recognition of visual clusters will be discussed in Chapter 7. The visual recognition and interpretation of more or less coherent patterns of graphical symbols in a matrix will be compared with a cartographic model on the prediction of visual clusters on graduated circles on maps. In this model the visual perception of clusters is considered to be based solely on some elementary physical characteristics of the graphical elements and their distribution across the map.

Chapter 8 concerns the higher levels of motor operations and cognitive information processing. It contains not only aspects of interactive construction of patterns, but also aspects of strategy in solving a graphical matrix. Definitive solutions to these tasks as well as those concerning intermediate configurations during the process of structuring, will be recorded and validated to more objective statistical measures.

Chapter 5 Discrimination and Sizing of Elementary Graphical Symbols

5.1 Abstract

Research on size judgement of elementary graphical symbols lacks clear findings as to the actual ratio in area between symbols that is required to make them perceptibly different. Weber already noted that the required increase in area of a stimulus is equal to a constant fraction of its original size (for instance see Engen, 1971).
In his theory on graphic information processing, Bertin claims that a size ratio of 1.12 is necessary for visible discrimination of two circular dots. In the experiments discussed in this chapter, the ratio factor of Bertin was tested for three different symbol shapes, dots, squares and bars. Results showed that, for these symbol shapes, a size ratio of 1.12 is more than sufficient for visual discrimination of two stimuli in a pair, and for sizing a larger number of stimuli in a small matrix. Whereas performance in this more complex sorting task reached a maximum at relatively slight differences between the stimuli, the subjectively experienced ease of executing this task increased as differences grew larger. For reasons of efficiency and ease it is, moreover strongly recommended to use differences between graphical elements that are a number of times greater than the required ones as a minimum. Finally, it was shown that the actually required ratio between graphical stimuli depends on the shape of the symbols used and the number of dimensions in which stimuli are different.

5.2 Introduction

An important potential advantage of the use of graphics in cartographical maps is their inherent ability to portray the essential information in a straightforward, easily interpreted way by using symbols of elementary shape, such as dots, wedges, squares and bars. On maps, these symbols often give numerical information on a specific object or geographical location. But, whereas the use of single graphical symbols is mainly restricted to thematical maps, this cartographic domain is only one of the possible applications. In a different format, the data table, not only can graphics be used to obtain the same type of link between an object (or, to some extent, a geographical location) and its value, but can also be used in this tabular format. While a graphic representation of values on a map is quite commonly accepted, in a table, on the other hand, values are still solely expressed numerically. In this study we will combine the data-table construction and the use of graphical symbols by projecting these symbols in the cells of a matrix or table, creating a graphical matrix, table or data base in which the graphical symbols replace the normally presented numbers.

Bertin (1981) postulates some important advantages of graphical symbols in a matrix compared to the common numerical tables. Imagine a replacement of numerical values by graphical elements where the size of the elements corresponds to the height of the original number. Within this graphical matrix it is possible to see at a glance where the higher values are located. To get the same kind of impression in a numerical matrix, the contents of cells have to be compared sequentially. Especially when two or more values have an equal number of digits, which decreases or even rules out the possibility of comparing these numerical values visually, pointing out the larger one will be a time-consuming task.

A crucial assumption underlying the use of graphics for the purpose of picking out the higher values is that subjects are able to detect and correctly discriminate between the projected symbols. Size differences therefore should at least be so large as to be perceptible but, for reasons of ease and efficiency, preferably larger. Distinguishing between symbols when differences between them are small, not only results in a larger number of errors and requires longer response times, but, in addition, the task itself rapidly becomes tiresome and subjects dissatisfied. So far, little empirical research has been conducted to test the size ratios that have to be used in this type of practical application. Directly usable clues have still to be derived on the absolute or relative size of a difference in area between symbols needed to make them discriminative. The purpose of a first series of experiments, discussed in this chapter, is therefore to gain more insight into the size differences of symbols of elementary graphical shapes that are required for a perceptible difference.

The representation of numbers by elementary graphical symbols has been mainly explored in four different areas of research, namely, (1) the work of the French geographer Bertin, (2) cartography, (3) psychophysics, and (4) exploratory data analysis.

First, a summary of research is given on this topic in these four areas. This short review is followed by a section on the outline of the design of the experiments and their results. In the last part of this chapter, results are discussed and compared with previous findings and some recommendations are made on the practically useful size ratios of graphical symbols.

5.3 Different approaches compared

Although the work of Bertin (1981, 1983) gives a very comprehensive and coherent body of theory on several aspects of the display of information in a graphical format and the processing of this graphical information, there is still a lack of empirical evidence. A large number of rules stated by Bertin still lack empirical verification. Cartographic and explorative statistic research, on the other hand, are characterized by an often too practical approximation to the problem. Testing method (ratio estimation versus magnitude estimation[1]), instructions, the stimulus range used and the presence and size of a reference or standard, vary considerably between examined cartographical studies and form an obstacle to general statements and rules.

The same practical approach by studies in the field of explorative statistics (with their often very creative graphical formats) also militates against the generalization of individual experimental findings. The approach in psychophysical studies, on the other hand, is mostly very theoretical. Although interested in the verification of certain more general psychophysical laws, the findings of these psychophysical studies are mostly restricted to a large number of specific laboratory-like conditions, which makes practical application very difficult to assess.

Bertin's Graphical Sign System

According to Bertin (1981) seven differential variables can be distinguished when using graphics on a map. These are respectively, shape, orientation, colour, texture, value (grey tone), size, and location (on the x-y plane). The actual choice of one or more of these variables is dependent on the kind of information that is to be displayed (see Chapter 4 and Chang, 1978). If the information is on a low, nominal measurement level, variables, such as shape, orientation or colour can be used. If the information is on an ordinal level, the texture or value variables (with categories of grey tones) can be used. If the information is on an even higher measurement level (interval

[1] In the case of ratio estimation the experimenter presents two or more stimuli and asks the subject to state the ratio between them. In magnitude estimation subjects must assign numbers to a series of stimuli when instructed to make the numbers proportional to the apparent magnitudes of the sensations produced.

or ratio) only value, size and location can be used correctly as variables of the image. An object that is twice as heavy as another object can be represented by a symbol that is twice as large (but not twice as "dark", as we have seen in Chapter 4).

Since the information of interest in the present study is mainly on a high measurement level and the projection of symbols in a matrix implies the impossibility of using location as a variable, attention is directed to the symbol size as a variable.

Thematic Cartography and Graphical Symbols

The use of graphics is almost inherent in the field of cartography. Even the elementary display of areas on a map makes use of shape as a graphical variable. Graphical variables in cartography are very widespread, ranging from the use of colour to indicate the difference between land and sea for example, or outstanding symbols to indicate country borders, to the use of variable sizes of arbitrary symbols to suggest different quantities. Cartographers have systematically studied the judgements of perception and size regarding some elementary graphical symbols, particularly over the past three decades. The circle or dot, in particular, has received close attention as a graphic representation of numerical information. The extent to which map readers can make quantitatively accurate comparisons of the size or area of graphical symbols, is emphasized in many of these studies (e.g. Dobson, 1974; Cox, 1976; Chang, 1977, 1978; Flannery, 1971; Meihoefer, 1969, 1973).

Chang (1980) notes three factors that have an influence on this accuracy of comparison. The first concerns the testing method, in which the accuracy of judgement depends on whether the method is one of magnitude estimation (e.g. Cleveland et al., 1982; Chang, 1977) or ratio estimation (e.g. Flannery, 1971; Crawford, 1971). The second factor relates to the instructions that are given to the subjects (e.g. is there emphasis on the area or the size of the symbol? Teghtsoonian, 1965) and the presence and size of a standard symbol as reference (e.g. MacMillan et al., 1974; Meihoefer, 1969, 1973; Cox, 1973). The third factor concerns the range of stimulus sizes and the sequence in which they are presented or their estimates are obtained.

It should be clear that almost all cartographical studies using graphical symbols direct their attention to the interpretation of the size of these symbols and not to their mere discrimination. The differences in size of the symbols actually used are therefore much greater than the minimum required for visual discrimination. Most of these studies conclude that the estimation of the size of a symbol compared to a reference symbol is more or less correct, given that absolute differences in size between the symbols are 'sufficiently large'. What is meant by sufficiently large unfortunately remains unclear.

Psychophysics

The first systematic research on detection, discrimination and size estimation of physical stimuli in the field of psychophysics actually dates from the last decades of the nineteenth century when Weber and Fechner were interested in determining the threshold values of sensation on a number of stimulus dimensions. Weber discovered that the amount of change in stimulus intensity (ΔI) that is necessary to allow detection of this change is approximately proportional (C: Constant) to the size or intensity of the standard (I). This finding has become known as Weber's Law, ($\Delta I/I=C$). Fechner saw in this law a means of measuring sensation quantitatively. His equation, derived from Weber's law, stated that the magnitude of a sensation (S) grows with the logarithm of the initiating stimulus (I); $S= c(\log I)$. The majority of psychophysical studies that followed emphasized such quantitative measurements of sensations (see the next chapter in the present thesis) and moved away from Weber's law, not least because constant Weber fractions were never observed throughout any stimulus dimension. At relatively low stimulus intensities (near the threshold of perception) and for some stimulus dimensions also at the upper end of the stimulus range, the Weber fraction was found to increase. The Weber fractions only remain more or less constant for various stimulus dimensions in the middle of the stimulus range.

Within psychophysical research there is therefore a very extensive range of experiments on the interpretation or estimation of stimulus size, which is comparable to the same type of experiments carried out within the area of thematic cartography. Research on stimulus discrimination, however, is mainly restricted to more theoretical investigations on the behaviour of Weber's constant and Fechner's Just Noticeable Difference (JND) or Difference Limen (DL) within different ranges of the stimulus (for a description of these and related ideas see e.g. Guilford, 1954; Torgerson, 1958; Engen, 1971; d'Amato, 1970). Most of these studies confirmed earlier findings that the constant increases when the absolute intensity of the stimulus decreases and approaches the threshold of perception. It turned out to be difficult to find practically useful indications on the required areal difference of dot symbols ranging from about one to a few millimetres in section.

Exploratory Data Analysis

Explorative statistics inquire into the applicability of a number of different kinds of graphs. Scatter plots, traditional pie and bar charts, and complicated plotting symbols such as Chernoff faces (Chernoff, 1973), glyphs (Anderson, 1960) and trees (Kleiner and Hartigan, 1980) have been tested for their specific advantages or merits (see also Chapter 3; Wainer & Thissen, 1981; Tukey, 1977 and Chambers et al, 1983). Simply shaped symbols (dots and squares) which are used in our study also belong to the range of investigated possibilities studied in the field of exploratory data analysis (Cleveland et al., 1983; Chambers et al., 1983). As with cartographical studies, however, the areal differences between graphical symbols that are used generally are clearly beyond the level of mere discrimination. The magnitude of minimally required differences have scarcely advanced further.

5.4 The discrimination of symbols on size

When comparing the results from these four sources, we note first that although much research has been done on the judgement of the size of graphical symbols in the fields of cartography, psychophysics and explorative statistics, very few studies give any clues to the absolute or relative size of a difference between symbols that is needed to make them discriminable.
The main question underlying many of the studies in the above-mentioned fields is one on the impression of size of a presented symbol compared to a reference symbol of known size. The question in this chapter however, is what the required difference (in length or area) between symbols has to be in order to make them discriminable. On practical considerations (the smaller the symbols, the more data can be presented in a matrix of a certain size) the main interest is on the 'just noticeable difference' in symbol sizes that are only a few millimetres in section. Whereas the conclusions of certain previously discussed studies are restricted to the claim that differences in size between symbols have to be 'sufficiently large', the

first series of experiments in this chapter investigates whether anything more specific can be said about the absolute size of this necessary difference.

A second inconvenience that is particularly conspicuous in the cartographical literature on the use of graphical symbols in maps, is the lack of a standard range of stimulus sizes. Different studies use very divergent ranges and within these ranges a different number of tested symbols (see Chang, 1980; MacMillan et al., 1974). As the aim of the present study is to test graphical symbols for future use in a rectangularly shaped matrix, it is, within this context of application, very important to reduce the absolute size of the largest symbol to a minimum. Diameters of several centimetres, quite common on cartographical maps, are very impractical when large data files are to be presented.

Bertin gives us some more hold on the actual difference in size that is needed to be detectable. When using circular dots of proportional size as quantities, Bertin (1981) states that, in order to provide a slight but still perceptible difference between two successive dots, the ratio between their areas must be approximately 1.12. With this ratio it is possible to discern as many as 20 steps that provide a perceptible difference between two dots A and B where the surface of A is 10 times that of B ($1.122^{20} \approx 10$). With a larger number of steps, the differences would become invisible according to Bertin. This ratio is not operative for the complete range of perceptible symbol sizes. When the absolute size of symbols becomes smaller, the differences need to be even larger to allow discrimination. Since very small dots require a ratio larger than 1.12 (Bertin, 1981 p. 207), the absolute size of dots that can be used is limited toward the bottom. Although not explicitly stated, we presume that these absolute sizes relate to a "normal" reading distance of some 30 centimetres.

So, in recommending an absolute ratio of about 1.12, Bertin is much more specific than the "sufficiently large" difference in size that was requested by some studies in the fields of cartography and explorative statistics.

Although the observation of a required constant ratio by Bertin is in accordance with Weber's Law ($\Delta I/I=C$) as is the increase in ratio when the absolute values become smaller, this specific factor of 1.12 has, as far as we know, not been empirically tested by Bertin.
The same lack of empirical evidence applies to the larger ratio factors that are claimed to be necessary with very small dots.

Three observations on the value of 1.12.

In the first place, this value is of approximately the same size as the exponents which are found in the estimation of the area of circles in the context of the power law (this theme will be discussed in Chapter 6 of the present thesis). The detection of differences in size, on one hand, and the estimation of the ratio between the size of a variable circle and a standard (reference) circle, on the other are, however, not directly connected. Starting from the principle of the power law and its relevant exponent, it is not possible without further preface, to calculate how large the just noticeable difference between two areas has to be.

Second, it should be noted that it is, at least, remarkable that, with an enlargement of an area by a factor of 10, there are exactly 20 steps that provide a perceptible difference. Wouldn't it be possible that this specific factor originates from a more practical point of view?

The third observation concerns error tolerance. In a visual discrimination task the proportion of incorrectly answered trials will gradually decrease as the differences between the stimuli become larger. Bertin neither mentions the minimum proportion of correct answers he deems necessary, nor the proportion of errors he expects will be made at a ratio of 1.12.

Considering the above, the following questions are pertinent:

1. Is a ratio of 1.12 between the area of two graphical symbols, as found by Bertin, sufficient for the discrimination between these symbols?
2. Does the human ability to discriminate between graphical objects of different size depend on the shape of the symbols used?

To test these questions, symbols of three different shapes were used in an experiment, namely, circles (dots), squares and bars. As the differences between two circles or squares are two-dimensional (area) whereas that between two bars is one-dimensional (height) the required difference between bars may be expected to be smaller than those between circles or squares. It remains to be seen, however, whether it is indeed easier to compare symbols differing in size in one dimension than in two or more.

In the first experiment the factor of 1.12 mentioned by Bertin was used as a fixed enlargement factor. The height of subsequent bars increased by this fixed enlargement factor whereas the diameter of circles and the side length of squares increased by the square root of this factor (an increase in area of a circle or square by a factor of n gives an increase in diameter or side length of $\sqrt{n}$).

The line of approach of the present study stands midway between the more practical approach, for example that followed in cartographical studies and the more theoretical approach that is found in psychophysical studies. In view of our interested in a clear, practical application of certain simple graphical symbols, a selected number of specific conditions, for example the testing method found useful in various cartographical studies, was used. In addition to these methodological considerations, the series of symbol sizes proposed by Bertin in his Graphics and Graphic Information Processing (1981), was chosen as the object of the study. Within these limits, the intention is to follow a conventional experimental method used in psychophysical research, as far as possible.

5.5 Experiment 1

The objective in the first, short paper-and-pencil experiment was to check whether the proposed ratio of 1.12 was sufficient to allow a correct discrimination between two stimuli presented pairwise. In this "pilot-like" experiment, three series of differently shaped symbols were used, dots (circular), squares and bars. Each series contained 24 symbols of different size. All the symbols were black. The ratio in surface area between two consecutive stimuli in a series was equal to 1.12, that proposed by Bertin. Stimuli were presented pairwise on paper cards with the underside of the symbols vertically aligned and a horizontal distance of 5 mm. between the centres.

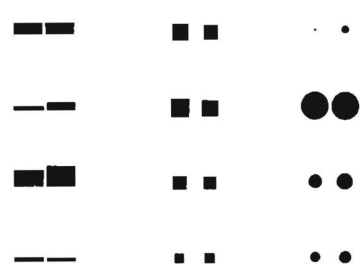


Figure 5.1 Examples of the stimulus material presented in experiment 1

Fifteen pairs of stimuli were drawn on each of the cards. Each of the symbols was compared to the nearest three larger and nearest three smaller symbols in the series, that is symbols that were 1.12, 1.12^2 ($\approx$ 1.25) and 1.12^3 ($\approx$ 1.40) as large or as small, as well as to a symbol of exactly the same size.

The surface area of the symbols ranged from about 1.4 mm^2 to about 19.6 mm^2, which corresponds to respectively about 15 minutes of arc and about 1 degree of arc at a reading distance of 30 cm.. These sizes correspond to stimulus sizes recommended by Bertin. A total of 22 students participated in this self-paced experiment. The test was performed by 6 subjects at the same time and took 20 to 30 minutes. All subjects judged all possible stimulus combinations with all three symbol shapes, a total of 270 trials. The total of 270 pairs of stimuli was distributed over 18 cards, 6 for each of the symbol shapes. Presentation of cards was counterbalanced across subjects. The task of the subjects was to determine the larger stimulus in each pair. For this purpose they had three response categories; right-hand stimulus is larger, left-hand stimulus is larger, stimuli are equal in size.

Table 5.1 Percentages of errors for each of the symbols and the ratios between them. W.S. indicates "wrong symbol" and gives the percentage of trials in which the smaller symbol was considered larger. Percentages: number of trials wrongly answered/total number of trials in this category.

ratio	**dot**			**square**			**bar**		
in area	W.S	F.A.	Miss	W.S	F.A.	Miss	W.S	F.A.	Miss
0.00	---	10.3	---	---	18.2	---	---	4.0	---
1.12	0.3	---	27.1	0.5	---	21.7	0.2	---	4.2
1.12^2	---	---	1.2	---	---	1.4	---	---	0.2
1.12^3	---	---	0.4	---	---	---	---	---	---

Results and discussion

For an areal ratio of adjacent stimuli of 1.12, the percentages of incorrectly answered trials were 27.4, 22.2 and 4.4 for, respectively dots, squares and bars[2] (see Table 5.1). For an areal ratio of 1.12^2 these percentages were, respectively 1.2, 1.4 and 0.2. Although the total number of incorrect responses for the ratio of 1.12 is rather large, this was probably due to the

[2]Experiments in which a relatively large number of subjects is used and in which an inclination towards averaging of results over subjects can be seen, are in contrast to experiments in which the number of subjects is restricted. As we are interested most of all in the practical usefulness of a graphical tool, a relatively large group of subjects is used for the different experiments. This does not mean that results of subjects or even of individual trials were added or averaged without due consideration. When only averages are displayed and no reference is made to individual differences, this means either that these individual differences were not significant or that they were not systematic.

presence of the response category "stimuli are of equal size". Subjects could deal with the experimental task easily by showing a tendency to answer that stimuli were equal in size.

In Table 5.1, it can be seen that 98.1% of all incorrect responses in trials in which the ratio in areas between stimuli was 1.12, fell into this category of "misses" or type-I errors ((27.1 + 21.7 + 4.2) / (0.3 + 27.1 + 0.5 + 21.7 + 0.2 + 4.2)). In the remaining 1.9% (100 - 98.1) the smaller symbol was considered the larger.

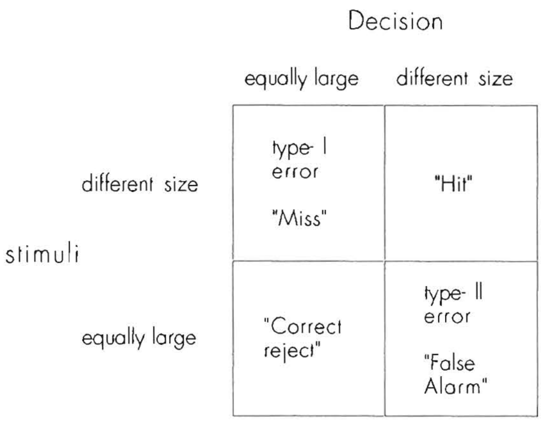


Figure 5.2 The four cases generated by combinations of two decisions with two actual situations. Errors of types I and II are identified in two cells of the table. In this presentation it holds that H0: stimuli are equal, and the alternative H1: stimuli are unequal.

Most of the errors in the experiment were of type I. The second largest source of errors was in trials where similar stimuli were judged as being of different size (type-II error or "False alarm"). Of the total number of trials where stimuli in a pair were of the same size, 10.8% was incorrectly answered. Only a small number of errors was made where the ratio was 1.12^2 or 1.12^3.

Type-I and type-II errors represent different response strategies. Whereas the error of seeing differences between symbols of precisely the same size (type II) indicates risk-taking behaviour (low criterion value), the error of not seeing differences between two symbols of unequal size (type I) are rather to be expected with subjects who show cautious response behaviour (high criterion value). For a discussion of response strategies and corresponding criterion values see Massaro, 1975. Some subjects judge two symbols only as being different when the difference is obvious to them.

These subjects often judge symbols to be of the same size when they actually differ slightly, thereby making a type-I error. The same subjects, however, will rarely make errors of type II. Other subjects only need a very slight difference before deciding that two symbols are not of the same size. These subjects make very few type-I errors but a relatively large number type-II errors.

Of the total of 22 subjects, only 4 were responsible for half of the type-II errors, whereas 6 others were responsible for the greater part of the type I errors. Errors made by the other 12 subjects were more evenly distributed [illegible] the two error categories.

The slight differences between the presented symbols can be expected to have an influence on the decision processes or the risk-taking behaviour of certain subjects and will be reflected in their criterion values. When these subjects become aware that the differences are very slight, it is very likely that the criterion on which they decide that one of the two symbols is larger, in fact becomes lower. This aspect has also to be considered when deciding what sizes and heights of symbols can be used in a graphical data matrix.

Gilmartin (1981) discusses the predisposition of a subject in an experimental context to respond in a particular manner or to attend selectively to some aspect of the stimulus. According to Gilmartin, this increases the probability of a certain kind of response which may introduce a bias in the results of an experiment. She warns that predispositions in response behaviour are not only created deliberately by the researcher as part of the experimental design, but that they can also occur unintentionally.

In addition to these two major causes of discrimination errors, momentary lapses in attention on the part of the subjects can also account for some of the errors made. It is very probable that this kind of error occurs equally often with stimulus pairs that have a very slight difference in size, thus demanding the same continuously high level of attention, as with stimuli of the same size.

In conclusion, the results of the first experiment indicate that even though a ratio of 1.12 seems to be on the low side, this can be due to a number of characteristics of the experimental design, in particular the introduction of the response category; "stimuli are of equal size". In a second experiment, therefore, this response category will be omitted, obliging subjects to make a more conscious and serious choice between one of the stimuli in a pairwise presentation.

5.6 Experiment 2

Stimulus material and environmental conditions

Stimuli were black symbols, presented on a white CRT screen. Three series of differently shaped symbols were used, circles (dots), squares and bars (columns). The reference stimuli (Ss) were presented alongside and had to be compared to a variable stimulus (Sv). Each series contained 7 different standards, varying in size from 9.8 mm. to 50 mm in diameter. The size of 9.8 mm. corresponds to about 11 minutes of arc at the applied distance between subject and screen of 3 metres and is about equal in visual arc to a symbol of 1 mm. in diameter at a reading distance of 30 cm. A size of 50 mm. corresponds to a visual angle of 1 degree of arc and is comparable to a stimulus of 5 mm. in diameter at reading distance. The horizontal heart-to-heart distance between the stimuli was 9.5 cm. (at the applied distance of 3 metres this corresponds to 1 degree 50 minutes of arc).

In about half the trials the Ss was presented to the left of the Sv, in the other half on the right-hand side. The position of the Ss was randomized. For dots and squares, the ratio in surface area between standard stimulus and variable stimulus ranged from 0.79 to 1.21 in steps of 0.03. For bars these ratios varied from 0.88 to 1.12.

Because a standard Macintosh 72-dots-per-inch screen was used, the subjects were seated at a distance of 3 metres from the screen. At this distance, differences in length and width of a stimulus of about 24 seconds of arc can be secured and dots do not appear jagged but have a smooth circular shape.

Due to this restricted resolution of the screen of 72 dots per inch not all ratios could be used with each of the standard sizes as dots and squares had to be perfectly circular, respectively square.

The room in which the experiment took place was completely shielded from daylight and illuminated exclusively by 2 spotlights and the computer screen itself. Conditions of illumination were measured before the experiment started and were constant during all the experimental sessions (the contrast between the white screen and the black stimuli was about 50:1).

Instruction

Before the actual experiment took place, subjects were informed about the experiment and the task they had to perform. "Compare the two stimuli that are presented pairwise in each trial. When the right-hand one is larger, answer "yes" by pressing the "j" key on the keyboard, otherwise answer "no" by pressing the "n" key". As soon as one of these buttons was pushed, the two stimuli on the screen were erased and the next two stimuli appeared. In that way, the experiment was self-paced and the duration of stimulus presentation was controlled by the subject. Incorrect responses could be corrected by pressing the f (false) button. The complete trial was redone, the stimuli of this erroneous trial were again presented on the screen. At each 25th trial subjects were informed as to the number of trials they had finished, each 100 trials there was a pause of 10 seconds and after the first and second session there was a short break of about 10 to 15 minutes. The instruction and a number of test trials took about 10 minutes, During this period, subjects could get used to the lighting conditions. After the instruction, but before the actual experiment started, the eyesight of the subjects was tested using a "Landolt" test (see section on subjects).

Experimental design

The experiment comprised three sessions. At each of the sessions 360 (bars) or 400 (dots and squares) pairs of stimuli were presented on the screen. Within a session all stimuli were of the same shape. Sessions were counterbalanced between subjects. The task of the subjects was to decide whether or not the stimulus on the right was larger than that on the left. The response itself consisted of pushing the j (yes, the right stimulus is larger than the left stimulus) or n (no, the stimulus on the right side is not larger than the stimulus on the left) key on the keyboard.

Subjects

Nine male subjects participated in the experiment, all students or staff members of the Faculty of Architecture, Building and Planning. Subjects were paid for participating. The complete experiment, including instruction, eyesight test and breaks took about two and a half to three hours per subject. Before the actual experiment started, the eyesight or corrected eyesight (some of the subjects wore glasses or contact lenses) of the subjects was tested using a "Landolt" test. All subjects had a visible acuity of 1.0 or better.

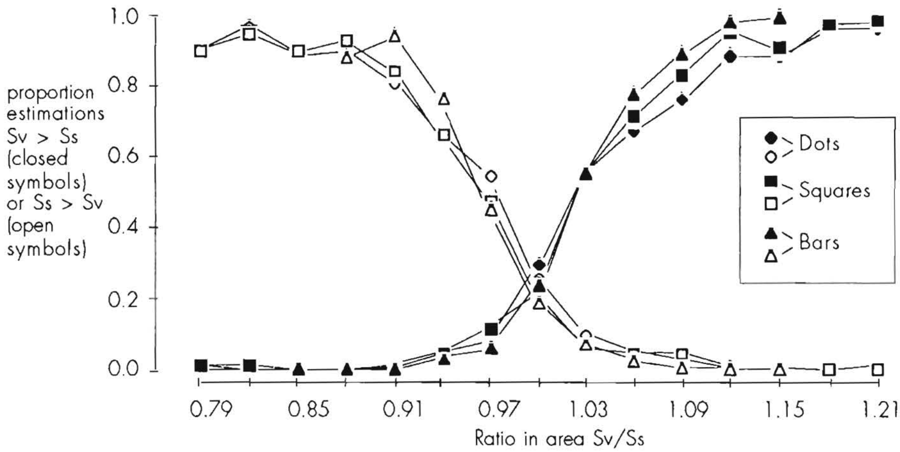


Figure 5.3. Proportion of trials in which the right-hand stimulus (Sv: closed symbols and Ss: open symbols) is estimated larger than the the left-hand one.

Results

In Figure 5.3 the total number of trials is divided according to the position of the Sv in the stimulus pair. On the x-axis, the ratio in areal surface between Sv and Ss is given. In the curves marked by the closed symbols, the Sv was presented to the right of the Ss. For ratios in areal surface Sv/Ss that are larger than 1.0 (the variable stimulus is larger than the reference) these curves show the percentages of correct responses or hits because the Sv was correctly judged larger than the Ss. For the Sv/Ss ratios of 1.0 and smaller, these curves show the percentage of incorrect responses. The Sv was equal to (ratio 1.0) or smaller than (ratios <1.0) the Ss, but was still judged larger. The responses by the subjects to these "closed symbol" trials were "yes, the right-hand stimulus is larger" whereas they should have been "no, the right-hand one is not larger".

The same description applies to the curves marked by the open symbols (the curves declining from left to right). These curves represent the trials in which the Sv was presented to the left of the Ss and responses were "yes, the right-hand stimulus is larger". For all ratios Sv/Ss that are smaller than 1.0 these curves show the hits. The standard stimulus was correctly judged larger than the variable stimulus. For ratios that are equal to or larger than 1.0 these curves again show the percentages of incorrect responses as, in these trials, the Ss was judged larger when in fact it was equal in size to the Sv (ratio 1.0) or smaller (<1.0). The two types of curves are nearly symmetrical around the x-value of 1.0. This means that the position of the Sv in the pair of stimuli appears to be insignificant and, what is more important, that subjects have no preference for the right-hand or left-hand stimulus.

It can furthermore be seen that the curves representing the bar symbols are steeper than those of dots and squares. Obviously, the discrimination of rectangular symbols is easier than that of circular and square symbols.

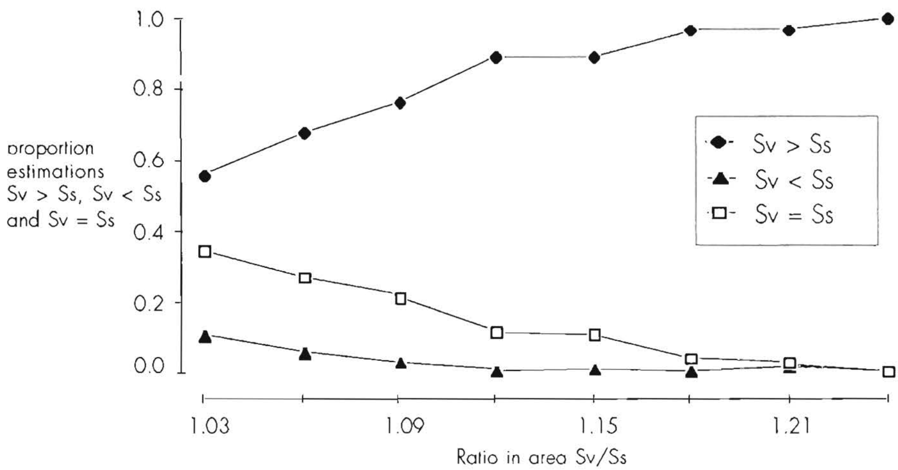


Figure 5.4. Proportion of trials in which respectively Sv (closed diamonds) or Ss (closed triangles) are judged to be the larger dot symbol in the pair. The percentage of trials in which both dots Sv and Ss are considered equally sized can be deduced from these curves (square symbols).

In Figure 5.4 the right-hand part of the dot symbol curve of Figure 5.3 is shown (that is, only Sv/Ss ratios larger than 1.0).
The top curve (marked by the closed diamonds) gives the proportion of trials in which the larger symbol was correctly judged larger. The bottom curve (indicated by the closed triangles) shows the proportion of trials in which the smallest symbol was considered larger. When we add up the percentages of both curves at each of the areal ratios (Sv/Ss) and subtract this total from 100% we have deduced the percentage of trials in which the symbols would be estimated as equally large, whereas one of them was in fact larger than the other (indicated by open square symbols in Figure 5.4). Such errors are the "misses" or type-I errors (see Figure 5.2).

As Figure 5.3 shows that there was no preference for left or right, and curves were symmetrical around the x value of 1.0, the proportions of type-I errors can be easily calculated for all ratios and for all three symbols. Connecting the proportions for all ratios (also those equal to and lower than 1.0) would give a curve that inclines towards the value of 1.0 and levels off when the difference in size between Ss and Sv increases.

An even distribution of these errors over the response categories (1), left-hand stimulus larger and (2) right-hand stimulus larger, in a two-alternative forced-choice task results in a standard "frequency-of-seeing function"; a graphical description of the changes in the perceiver's detection response as a function of the physical variable being manipulated in the experiment (Haber & Hershenson, 1980). This drawing is shown in Figure 5.5. Presentation of results in this way allows for a calculation of the D.L. (Difference Limen: minimum amount of stimulus change required to produce a sensation difference 50% of the time). By linear interpolation, for instance, we get ratio values of 0.038, 0.038 and 0.036 respectively for dots, squares and bars.

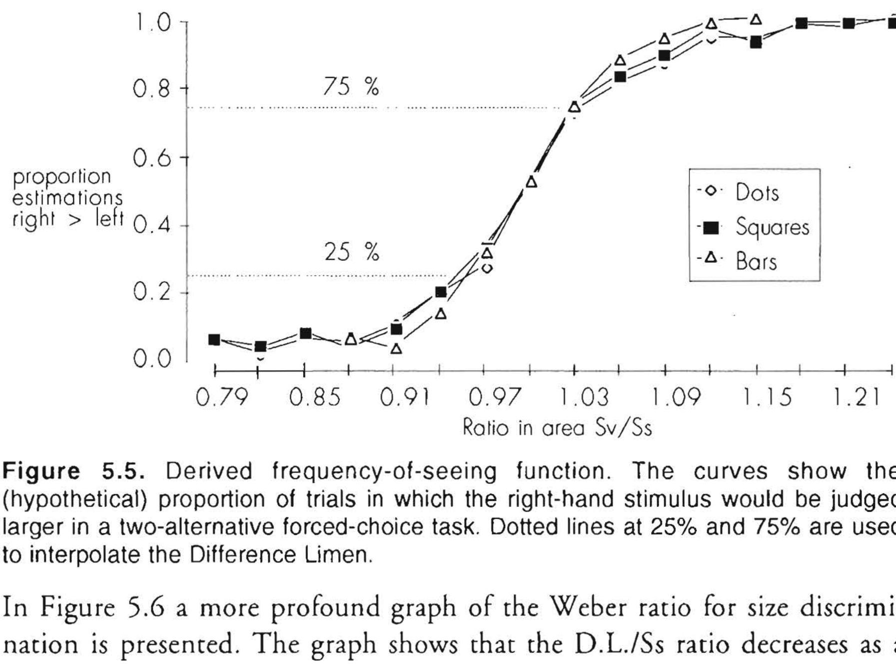


Figure 5.5. Derived frequency-of-seeing function. The curves show the (hypothetical) proportion of trials in which the right-hand stimulus would be judged larger in a two-alternative forced-choice task. Dotted lines at 25% and 75% are used to interpolate the Difference Limen.

In Figure 5.6 a more profound graph of the Weber ratio for size discrimination is presented. The graph shows that the D.L./Ss ratio decreases as a function of the magnitude of the standard stimulus. In this figure the Difference Limen is defined as

$$D.L. = (U.L. - L.L.) / 2 \qquad [5.1]$$

where the Upper Limen (U.L.) is the performance at the 75% level in a two-alternative forced-choice task and the Lower Limen (L.L.) at the 25% level.

For the dot and square symbols, separate curves are drawn for both increase in area of the standard stimulus and increase in diameter (dot) or side length (square). The D.L. curves for area show that the values for bars are lower than those of dots and squares, which indicates that differences between bars are easier to detect than those between dots and squares.

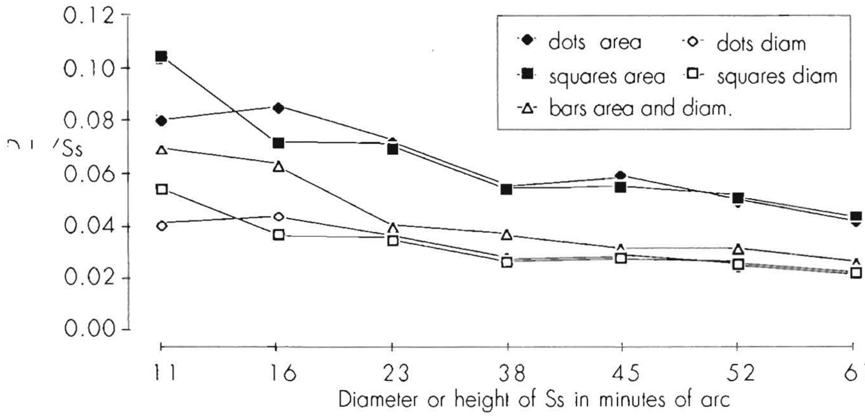


Figure 5.6. Weber ratio for size discrimination. The Difference Limen is the minimum amount of stimulus change required to produce a sensation difference in 50% of the trials.

Although overall differences in performance between subjects were not significant, we can note some differences in decision strategies between subjects corresponding to the cautious and risk-taking behaviour already mentioned in the discussion of experiment 1.

Subject A.G for instance was very careful in judging Sv larger than Ss. He tended to respond "no, the right-hand symbol is not larger" which he actually did in 62% of all trials. This cautious behaviour results, in fact, in a relatively small number of false alarm trials but also in a small number of hit trials. A strategy that is more or less opposite to that of A.G. was shown by subject G.H. This subject is more inclined to respond "yes, the right-hand symbol is the larger one", which results in more correct responses (hits) when the right-hand symbol is indeed the larger one of the two, but results in more incorrect responses when, in fact, both symbols are of equal size (false alarms) or the left-hand symbol is the larger one. Subject G.H. responded "no" in 55% of the trials.

Whereas the decision behaviours of these two subjects differed considerably, they resulted in a comparable overall performance. The rates of hits and false alarms for each of the subjects at an Sv/Ss ratio level of 1.03 are shown in the receiver operating characteristic (R.O.C.) plot in Figure 5.7. Even though subjects differed in their decision strategies, all trade-off points are close together.

Concluding the results of this experiment, we note that even though Bertin's ratio factor of 1.12 appears not be based on the Weber ratio (a D.L. of about 0.03 to 0.07 was found, see Figure 5.6), it is shown to be a very acceptable ratio between areas of two juxtaposed symbols, especially when high performance is required (a ratio of 1.12 resulted in a proportion correctly answered trials of about 0.9, see Figure 5.5).
It turned out that performance, as based on the Difference Limen, increased still more at symbol sizes of 1 degree of visual arc and beyond.
Bar symbols are easier to discriminate than circular dots and squares.

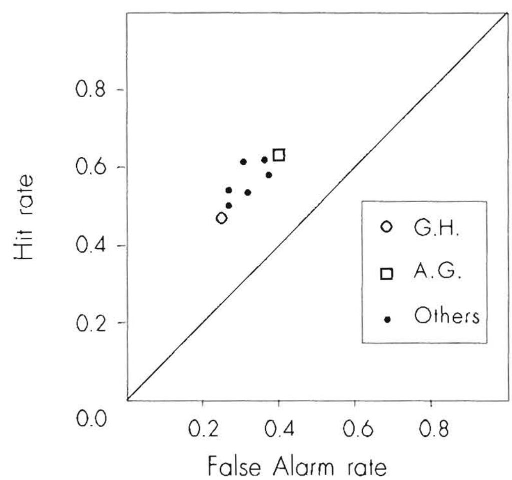


Figure 5.7. Receiver operating characteristic plot, resulting from plotting the hit rates and false-alarm rates for each of the subjects at a Sv/Ss ratio of 1.03. Absolute performance increases according as the plotted points draw closer to the upper-left-hand corner of the figure. Due to a disproportionate distribution of combinations of the Sv/Ss=1.0 ratio (required in calculating the F.A. rate) and the size of the reference stimuli only the Sv/Ss ratio of 1.03 could be plotted.

Comparison of experiments 1 and 2

The results of both experiments showed that, when comparing graphical symbols, the discrimination of two differently sized and juxtaposed bar symbols appears to be easier than the discrimination of dot or square symbols. In a total of 21.7% (squares) and 27.1% (dots) of the trials in experiment 1 in which the ratio in size between standard and variable stimulus was 1.12, subjects actually responded that these stimuli were of the same size ("misses"). A proportional allocation of these percentages to the two categories of a two-alternative forced-choice task results in, respectively 10.85 and 13.55 percent incorrectly answered trials. To these we have to add the 0.3 and 0.5 percent of the trials in which the smaller symbol was estimated to be the larger one (see Table 5.1). This results in a total of, respectively 13.85% and 11.35% incorrectly answered trials for, respectively circles and squares. Changing the response categories and forcing the subjects to make a more weighted response (two-alternative forced choice instead of a three-alternative forced choice) resulted in a decrease in these percentages to about 10 (Figure 5.5). Due to some differences in performance between subjects, these percentages do not differ significantly from the results of the second experiment (for both circles and squares p>.10). In addition, it is at least doubtful whether it is correct to divide the misses equally between the two alternative categories in a two-alternative forced-choice experiment (Engen, 1971, p. 30; Guilford, 1954, p139-142). This decrease could be due to differences in the experimental design of the two experiments. Furthermore, the results of the second experiment allow weighting a required percentage of correctly answered trials to obtain a ratio in size between two juxtaposed graphical symbols.

In the second experiment it was not possible to use each of the size ratios with each of the sizes of the standard stimulus, as already mentioned in the section on the design of the experiment. A more detailed analysis of the specific size ratios at each of the standard sizes showed the same tendency for all standard stimuli: performance clearly improved when the difference between the two stimuli in a pair increased. The best results, however, were obtained by the largest standard sizes, as could be seen in the Weber curves in Figure 5.6.

5.7 Experiment 3

In a third experiment the interest was in the visual discrimination of graphical stimuli in a more complex sorting task. In a reorderable graphical matrix, normally a large number of stimuli are presented that have to be visually compared and sized before it is at all possible to put similar stimuli into a group or cluster.

A second point of interest in this experiment was the subjectively experienced complexity (or ease) in completing the sorting task. An equal performance on two different tasks does not necessarily imply that both tasks are experienced as equally difficult and, in practical situations, where a

large number of visual comparisons have to be made, differences in symbol size that are only just large enough for a correct discrimination, might prove to be tiring in the course of time and eventually lead to a deterioration in performance or to a dislike of doing the task.
By a selection of differences between stimuli that are a number of times greater than the minimum required, the occurrence of these problems could be prevented.

Stimulus material

Three different sets of symbols were used as stimulus material: circular dots, squares and bars. All symbols were solid black. Stimuli were presented on a black-on-white standard Apple Macintosh 72-dots-per-inch computer screen. In each of the trials, 6 stimuli were presented in a 2 (rows) · 3 (columns) format (see Figure 5.8). This group of 6 stimuli consisted of three different pairs, with each pair containing two identical stimuli. The ratio in area between the largest stimuli and the middle stimuli was equal to the ratio between the stimuli in the middlemost pair and smallest stimuli.

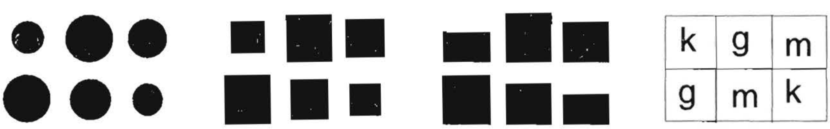


Figure 5.8. Layout of the stimuli as presented on the computer screen (left-hand part) and matrix with feedback on the responses that the subjects made by pressing a sequence of keys on the keyboard (right-hand part).

Actual ratios in area that were used were 1.06, 1.09, 1.12 and 1.15. The position of each of the 6 symbols in the 2 · 3 matrix was randomized. A total of 4 different sizes of the reference (middlemost) stimulus was used. With subjects seated at 3.5 metres from the computer screen, the stimulus sizes corresponded to respectively about 25, 35, 42 and 48 minutes of visual arc. The horizontal and vertical distance between the centres of the stimuli was about 63 minutes of visual arc. Height of the standard bars was 1.5 times its width. Lighting conditions were the same as in experiment 2.

Instruction

Before the actual experiment started, subjects were informed about the experiment and the tasks they had to perform.

The six stimuli had first to be ordered according to size, from small to large. This task was to be completed by pressing a sequence of keys on a 3 · 2 key pad. The layout of the keys of the key pad corresponded to the layout of the stimuli on the screen. Subjects received feedback on the responses they made; for the first two key presses the character k (klein: small) appeared on the screen at the selected positions, with the next two key presses the character m (midden: middle) was drawn and with the last two responses the character g (groot: large) filled the last two cells of the 3 · 2 matrix (see Figure 5.8).

The presentation of stimuli on the screen lasted 4 seconds, after which period the screen was wiped, a 3 · 2 empty matrix was drawn and subjects were allowed to type their responses. Responses during the presentation of the stimuli were not accepted and this was indicated by a short tone whenever that was the case.

After sizing the stimuli, subjects had to give their opinion as to the complexity of the sizing task by answering "it was very difficult to sort the stimuli" or "it was very easy to sort the stimuli" on a 9-point scale ranging from 1 (difficult) to 9 (easy). Before the answers to each part of the response (sizing and opinion on complexity) were actually recorded they had to be confirmed by pressing a key on the key pad. Therefore, as long as a specific response was not confirmed, it was possible to make corrections during the trial. Trials in which subjects responded too early and trials in which corrections were made were recorded.

Design

Each subject had to respond in 960 trials in all, distributed over two sessions. Each session contained three sections of 160 trials each. In each of the sections all stimuli were of the same shape. The sequence of the three sections was counterbalanced within the session, trials were randomized within sections. Every 10th trial the trial number was shown on the screen. After every 40 trials there was a pause of 1 minute, and after every section of 160 trials, a break of a few minutes.

The time interval between the two sessions was at least 4 hours and at most 2 days. The complete experiment took 3 to 4 hours.

Environmental conditions and tests of the eyesight of subjects were the same as described in experiment 2.

Subjects

Six subjects (three male, three female) participated, all students of the faculty of Architecture, Building and Planning. The complete experiment took 3 to 4 hours. Subjects were paid for participating. Eyesight of subjects was tested using the Landolt test. All subjects had a visible acuity of 1.0 or better.

Results

Performance

A four-way analysis of variance (with factors: subjects, session, symbol shape and ratio in size between Ss and Sv) was accomplished on the average performance (performance was averaged over 10 identical trials in each of the sections). This analysis showed a significant effect for all main factors and for the interaction between ratio and shape of the symbol (see Table 5.2). Although the differences in performance between subjects were significant (overall performance ranged from 47.1 % correct for subject 6 to 85.3% correct for subject 5) none of the interactions with the factor subjects was significant (all $p>.08$).

Table 5.2. Results of a four-way analysis of variance on the average performance (over 10 identical trials) of sizing 6 stimuli in a small matrix

Factor	d.f	F value	p value
Subjects (A)	5	83.49	.0001 *
Session (B)	1	20.60	.0001 *
AB	5	00.64	.66
Symbol shape (C)	2	68.16	.0001 *
AC	10	1.42	.17
BC	2	0.54	.58
ABC	10	0.53	.87
Ratio (D)	3	228.23	.0001 *
AD	15	1.61	.08
BD	3	0.58	.62
ABD	15	1.08	.37
CD	6	9.09	.0001 *
ACD	30	0.74	.83
BCD	6	0.59	.73
ABCD	30	0.74	.84

*: high level of significance

This means that the distribution of performance values across combinations of symbol shape and ratio levels were the same for all subjects. As we were mainly interested in the relative performance on the different symbol shapes, ratios and sizes of the standard stimulus, and because differences in absolute performance between subjects were significantly large, the average performances were standardized[3] (converted into standard z-values) for each of the subjects and the factor "subjects" was no longer considered in subsequent analyses (the z-values were treated as repeated measures).

The same procedure applied to the factor "session". Although overall performances were better in the second session, the increase in performance was found to be relatively the same for all subjects, ratio levels, symbol shapes and combinations of the three (all p>.37). Performances were thereupon averaged over the two sessions and the factor "session" was no longer considered in subsequent analyses.

Next, a three-way analysis of variance (symbol, ratio, size of the Ss) was performed on the z-values obtained (Table 5.3). This analysis showed a significant effect for all main factors and for the interactions between symbol and ratio and between ratio and size of the Ss.

Table 5.3. Results of a three-way analysis of variance on the z-values of performance in sizing 6 stimuli in a small matrix.

Factor	d.f	F value	p value
Symbol shape (A)	2	100.15	.0001 *
Ratio (B)	3	335.28	.0001 *
AB	6	13.71	.0001 *
Size of reference, Ss (C)	3	65.53	.0001 *
AC	6	1.50	.18
BC	9	4.71	.0001 *
ABC	18	1.42	.12

In Figure 5.9, performance in standard values (y-axis) is plotted to the ratio in size between Sv and Ss (x-axis). Each curve shows the results for one of the symbol shapes. This figure shows that the significant interaction between the factors symbol shape and ratio is largely due to the combination of high ratios and the bar symbols. The bar symbol shows better performance accompanied by a more profound "levelling off" than performance on dots and squares. Differences between bars, on the one hand, and dots and squares on the other, were significant at all values of the ratio (all p<.05).

[3]Analysing the results of this experiment we have the options of standardizing the original data (within subjects) into z-values before averaging across subjects, or directly averaging the original values. We have chosen the first alternative in order to be able to compare the normal distribution of errors (across the independent variables) between the subjects. Additional analyses, using the second option, revealed that the results of the two methods did not produce significant differences.

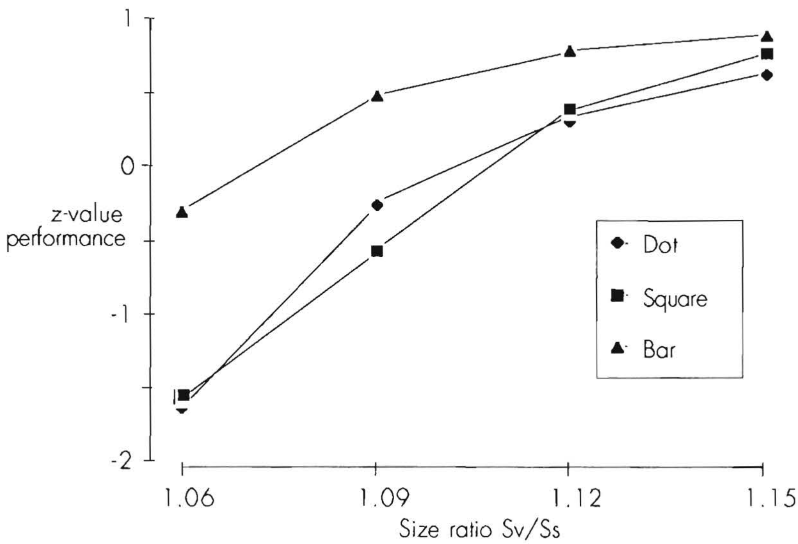


Figure 5.9. Performance (in z-values on the y-axis) for the three different symbol shapes, plotted to the different ratios in size between standard stimulus and variable stimulus

In Figure 5.10, size of the standard stimulus (in minutes of arc on the x-axis) is plotted to performance (in z-value on the y-axis) for each of the ratio levels. In this figure, performance is averaged over the three shapes of the symbols. Some distinctive features can be seen. Whereas performance is generally better for the larger stimuli, since all four curves incline towards the right, it also shows that differences in performance between ratios decrease at greater standards (differences in y-values drop).

Furthermore, a greater increase in performance at lower ratio levels than at higher ones can be seen since the two bottom curves (curves marked by the open and closed diamonds) show a greater increase toward the right than the two top curves (open and closed square marks).

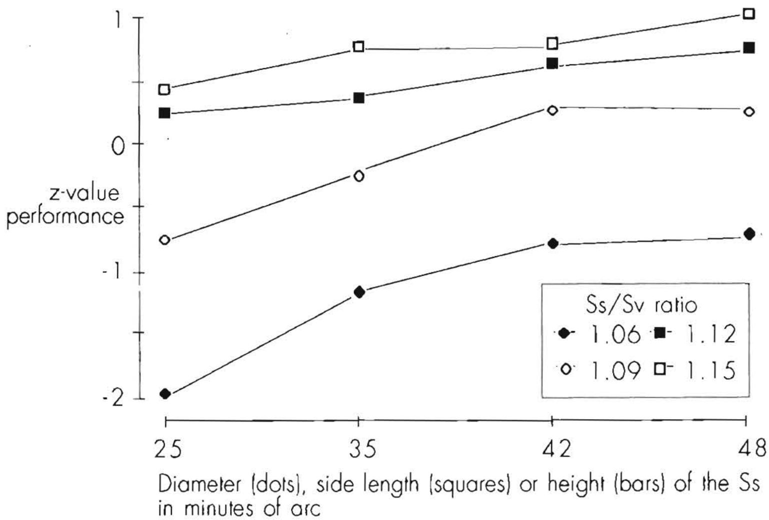


Figure 5.10. Performance on the sizing task (in z-value on the y-axis) plotted to the size of the standard stimulus in minutes of arc (on the x-axis). Each of the curves represents the performance for a single Ss/Sv ratio level at different standard sizes. Results are averages over dots, squares and bars.

Table 5.4. Results of a four-way analysis of variance on the average judgement (over 10 identical trials) of sizing 6 stimuli in a small matrix

Factor	d.f	F value	p value	
Subjects (A)	5	479.85	.0001	*
Session (B)	1	4.67	.03	*
AB	5	4.18	.001	*
Symbol shape (C)	2	85.19	.0001	*
AC	10	5.13	.0001	*
BC	2	0.54	.58	
ABC	10	4.16	.0001	*
Ratio (D)	3	155.52	.0001	*
AD	15	3.30	.0001	*
BD	3	0.45	.71	
ABD	15	0.70	.78	
CD	6	1.79	.10	
ACD	30	0.58	.96	
BCD	6	1.60	.15	
ABCD	30	0.83	.69	

Judgement

A four-way analysis of variance on the average judgement of the factors subjects, session, symbol shape and ratio, showed a significance for all main factors and for all second-order interactions involving the factor "subjects"

(Table 5.4). This means that subjects tended to differ in their judgements for the two sessions, the three symbol shapes and the four levels of ratio.

The interaction between subjects and session was due to a differential judgement of the subjects for the two sessions. Some showed, on average, a lower judgement for the trials in the second session, whereas others judged trials to be easier or equally difficult in the second session.

Although the interaction between subjects and symbol shape was also found to be significant, all subjects clearly showed the same tendency in their judgements on sizing symbols of different shapes. Whereas dots and squares were considered about equally difficult to order, all subjects clearly found it easier to size bars.

The interaction between subjects and ratio levels was due to differences in the steepness of the judgement curves (plotting judgement to size ratio). Although the slopes of the curves differed between subjects, they all judged the trials to be easier according as ratios between Ss and Sv increased.

Since subjects showed the same tendency in their judgements, these response values were first standardized (converted to z-scores) for each of the subjects and treated as repetitions in addition.

As none of the interactions of session (except the interaction with subjects) was found to be significant, judgements were also averaged over sessions.

A three-way analysis of variance on z-values of the judgement showed a significant effect for the three main factors and for the interaction between symbol shape and ratio level (Table 5.5).

Table 5.5. Results of a three-way analysis of variance on the z-values of judgements on the sizing of 6 stimuli in a small matrix

Factor	d.f	F value	p value
Symbol (A)	2	146.63	.0001 *
Ratio (B)	3	227.89	.0001 *
AB	6	3.01	.007 *
Size of standard (C)	3	97.04	.0001 *
AC	6	0.67	.67
BC	9	1.93	.19
ABC	18	0.56	.92

This interaction is shown in Figure 5.11. Differences between bars and dots/squares were significant at all ratio levels ($p<.01$) while the difference between dots and squares was only significant at a ratio of 1.12 ($p<.05$; for all other pairwise comparisons between dots and squares $p>.05$).

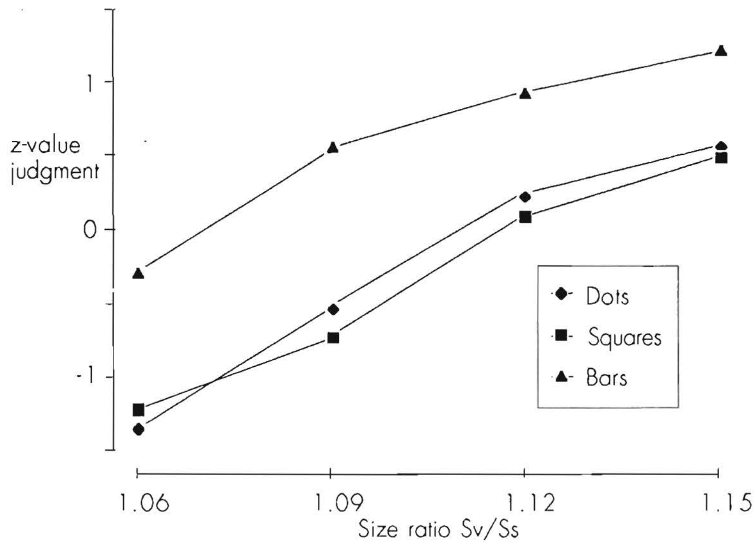


Figure 5.11. Interaction between ratio level and symbol shape for mean judgement on task complexity. Each of the three curves show the judgements for a specific symbol at various ratio levels.

Correlation coefficients between the z-values of the two dependent variables, performance and judgement and two of the independent variables, ratio level and size of the standard, are given in Table 5.6. Correlation coefficients of dots and squares are much alike, with the ratio factor showing a strong correlation with performance (about 70% of the variance in performance is explained by this factor) while the influence of standard "size" on performance is much less (size of the Ss only explains 7% to 11% of the variance in performance)[4].

Regarding the judgement values on dots and squares, the size of the standard appears to have a greater influence, while that of the ratio level happens to decrease. For bars, the correlation coefficients between independent variables and performance and judgement are less divergent, both factors have a more or less similar explanatory value when it comes to prediction of performance or judgement.

[4]In Figure 5.14 it can be seen that actual relations between ratio and performance or judgment are not linear because the drawn functions show a levelling of performance and judgment values with larger ratio factors. A quadratic equation would therefore result in a better fit of the actual relation. As values in Table 5.6 are based on a linear equation, these values are underratings of actual correlations between independent (ratio factor) and dependent (performance and judgment) variables (Cohen and Cohen, 1983, p.224-230).

There is, furthermore, a strong correlation between performance and judgement. The judgement values seem to be a fairly reliable reflection of expected performance.

Table 5.6. Correlation coefficients between the independent variables, ratio and size of the standard, the dependent variables, performance and judgement and mutual correlations between performance and judgement. Coefficients are given for each of the symbol shapes.

Dependent variables	Ratio	Ss	Perf. Dots	Perf. Squares	Perf. Bars
Performance dots	0.83	0.26			
Performance squares	0.84	0.33			
Performance bars	0.61	0.50			
Judgement dots	0.78	0.47	0.86		
Judgement squares	0.77	0.44		0.86	
Judgement bars	0.63	0.52			0.76

Discussion of experiment 3

Both performance and judgement on the ease of executing a complex sorting task show that Bertin's recommended ratio of 1.12 can be considered sufficiently large for sizing simple shaped graphical symbols. As performance reaches its maximum value of 100% at higher ratios (see Figure 5.9), the performance curves level off and differences between the three symbol shapes more or less disappear. Judgement on the ease of the task, however, keeps improving with big differences between stimuli, and bars are still considered easier to size than dots and squares. Therefore, greater differences than are necessary from the performance point of view, besides using bars as symbols in the matrix, could be recommended.

In addition to the ratio in size of the stimuli used, their absolute size also appears to have an effect on both performance and ease of sorting. As regards the sorting of stimuli, the recommendation is to use larger stimuli, especially when the differences between the stimuli are slight.

Correlation coefficients between dependent and independent variables are very high. Performance as well as judgement can be predicted very accurately (particularly with dots and squares), when only ratio level and size of the standard are known.

5.8 Conclusion

Results of all three experiments on the discrimination of two or more graphical symbols show a better performance for bar-shaped symbols than for circular dots or squares. In addition, subjects also judged the sorting task to be easier when bars had to be discriminated. Differences between bars on one hand, and dots and squares on the other are twofold.

- First, a change in the size of a bar is reflected in only one of the dimensions of the symbol (the other dimension is fixed). With dots and squares, changes in size are evenly distributed over both dimensions. Therefore the relative change in the variable dimension of a bar is a larger one, compared to unidimensional changes of dots (diameter) and squares (side length).
- Second, and closely related to the first observation, the difference in area between bar symbols is accompanied by a difference in shape between the symbols; one of the bars in the pair of stimuli is relatively more elongated or in the opposite case, thicker. With dots and squares the shape remains unchanged when there are changes in areal size.

The second explanation, however, is not sufficient to explain the differences found in performance between bars and dots/squares. Although two differently sized bars might be easier to know apart, their mere discrimination does not yet indicate which of the two symbols is the larger. In order to point out the largest or smallest symbol, heights or height-width ratios still have to be compared and evaluated.

The second explanation of unidimensional change for differences in performance between bar symbols and dot symbols, is illustrated in Figure 5.12. On the x-axis, the differences in height (not area) between the standard and variable stimuli are shown and performance (or performance corrected for chance) is given on the y-axis. For bar symbols increase in area is linearly related to increase in height, for dot symbols an increase in area equals the squared increase in height.

Trials in which dot symbols had to be sized show better performance than the bar trials. Obviously, subjects not only look at the pertinent difference in height between dot shaped symbols but also consider other aspects (e.g. size), when ordering these symbols.

The value of 1.12, proposed by Bertin, has not been derived directly from the Weber ratio, as this last is much smaller for a visual discrimination of two symbols of elementary shape. Results of the experiments described in this chapter, however, show that the value of 1.12 is a very reasonable one. First, a ratio of this value results in good performances in ordering tasks, as it is a number of times greater than the Weber ratio. Second, it allows for tasks that are simple and easy to perform.

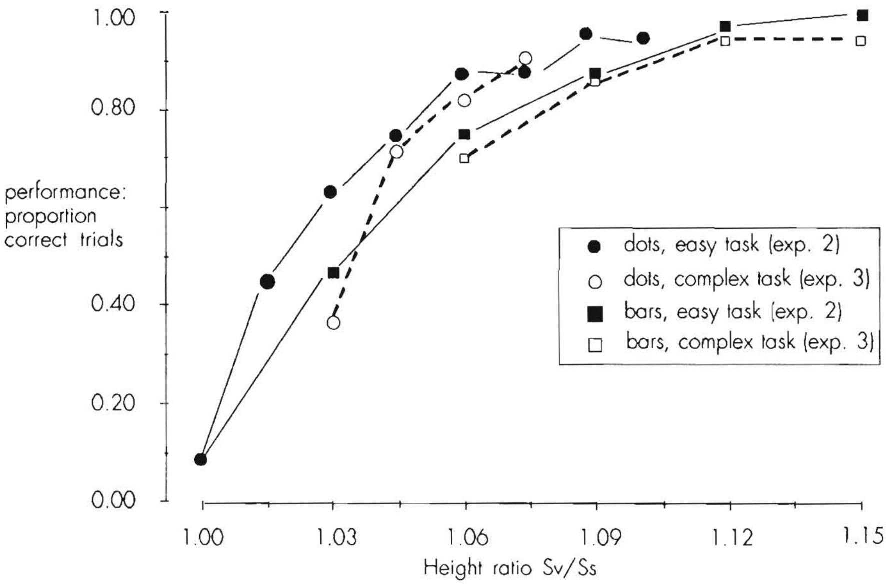


Figure 5.12. Comparison of the performance in a complex sizing task (sizing 6 graphical elements in a 3·2 matrix, exp. 3) and corrected performance in an easy sizing task (a two-alternative forced choice visual discrimination task, exp. 2). Actual performance on the y-axis (see also explanation of Figure 5.13) plotted to ratios in size between standard and variable stimuli.

A second contribution in our check on the proposed ratio of Bertin is a trade-off between performance and differences in size. The second and third experiments showed a smooth, continuous function between these two factors. Depending on the proportion of errors that are allowed in a particular experimental context, an optimum difference in ratio between graphical elements can be determined for each of the three tested symbol shapes. In the first instance, results of both experiments are not completely comparable. When only two symbols have to be compared, pointing out the larger one by chance is 1/2. Selecting the right order of 6 elements (3 pairs of 2 identical stimuli) by chance is only 1/90[5]. This chance factor can be

[5] With two correct alternatives out of 6 (because the two smallest symbols had to be pointed out in the first and second guess, only the selection of these two result in a correct answer), the chance of selecting a correct one at the first guess is 1/3. The chance of subsequently selecting the second correct alternative, with 5 items left, is 1/5. For the two middle-sized stimuli that had to be pointed out next, the chances become respectively 1/2 and 1/3. Now only the two largest symbols are left, so these are automatically selected correctly. Selection of all 6 items in the correct order becomes 1/3 · 1/5 · 1/2 · 1/3 ·1/1 ·1/1 = 1/90.

ruled out by a correction of the results of the simple task. Original and corrected performance for dot symbols on the easy task and performance on the complex sizing task are compared in Figure 5.13. Differences only occur at a ratio level of 1.06, at higher ratios both tasks are equally difficult to perform.

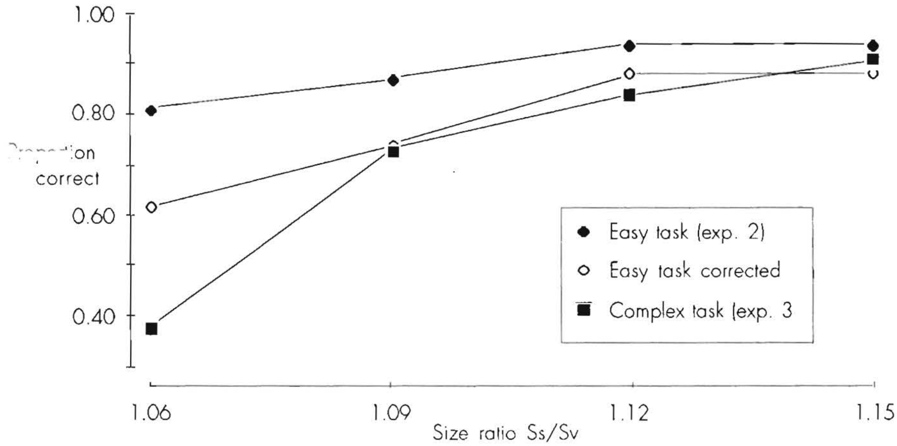


Figure 5.13. Comparison of the performance for dot symbols in experiment 2 (a two-alternative forced choice visual discrimination task) and experiment 3 (sizing of 6 graphical elements in a 3·2 matrix)

5.9 Further Research

Since error rates with symbol discrimination showed a dependence on the shape or dimension of the symbol, it would be interesting to devote more research to this relation between just noticeable visual differences and the dimension of this difference (length, area, content).

That such a table has a far-from-neglible practical value is shown, for example in a study by Vroon (1978). In this study it was experimentally confirmed that too small a difference in size between a current coin and a new coin that was brought into circulation in the Netherlands caused a lot of confusion. Experimental results showed that in about 15% of the presentations, one of the coins was mistaken for the other. The practical consequence of the introduction of the new coin was that within a short time the new coins appeared all over the country with small round stickers glued to them.

Chapter 6: Visual estimation of size of some graphical symbols in a matrix

6.1 Abstract

Although graphical symbols are often found to be a strong visual aid in the presentation of numerical information, the actual use of individual graphical elements is almost entirely restricted to cartographical maps. This chapter investigates the possibility of extending their use to a matrix context which means that location and meaning of the presented graphical symbols are uncoupled. By analogy with a number of discussed cartographical and psychophysical studies, some differently shaped graphical symbols are presented in a matrix in our experiment. Subjects had to estimate the relative sizes of these symbols and base their estimations on a small number of reference symbols.

Results show that the size of bars is estimated fairly accurately. Estimation of the size of squares and circular dots presented in a matrix shows an average overrating of actual symbol size. Exponents of power functions (between the logarithm of the actual stimulus size and the logarithm of the estimated size) were very close to 1.0. The use of grid lines separating rows and columns in the matrix had an ambiguous, detrimental effect on the accuracy of estimation and should therefore be abandoned as far as judgement of size is concerned.

6.2 Introduction

Although the use of graphical symbols for presenting numerical information has come in for much attention and research in the last three decennia, this phenomenon has, for the greater part, remained restricted to cartographical applications. Only recently has awareness grown that the visual strength of elementary symbols, such as circles and squares can also be applied beyond the context of cartographical maps. Cleveland, Harris and McGill (1982, 1983) for instance showed that circles could also be used on statistical maps where location of the presented symbols is of minor importance. Bertin (1981) even went further in proposing to project little graphical elements into the cells of a table-shaped matrix. Here a certain point symbol is marked with the labels of the row and column in which it is located, but the position of the symbol within the matrix is quite arbitrary.

Because the use of graphical symbols as advocated by Bertin differs in both layout and purpose from the well-known cartographical applications, it is necessary to make some further inquiries into the different stages of this transition. As for Bertin, an imperfection in the execution of his theory can be perceived at this point because his experiences with the use of graphics in a cartographical environment are too easily transferred to the different context of a graphical matrix. Therefore additional cartographical and psychophysical research on this subject will be studied in order to make good the deficiencies in Bertin's theory. Whereas cartographical research on the use of graphical symbols on a map is mainly performed in order to warrant a certain course of action, psychophysical research on graphics is primarily interested in the verification of certain more general psychophysical laws.

A first requirement in the use of graphical point symbols was discussed in Chapter 5. It was shown that a certain ratio in size between pairwise presented symbols is necessary in order to be able to discriminate between them. With a size ratio of 1.12, as was proposed by Bertin, a correct discrimination in about 90% of the stimulus presentations can be warranted. In this chapter, a second requirement, the correct estimation of the intended numerical value of graphical symbols will be investigated. In Chapter 4 it was shown that data of an interval or ratio type can only be correctly displayed graphically with the variables of size or, to a lesser extent, of value (grey tone). Therefore, the "size" variable seems to be the best one to use when it comes to estimation of intended numerical value.

In the following sections some practical implications of the translation of numerical values into the size of graphical point symbols will be discussed. First of all, the graphical elements must be visually discriminable, as we already noted. Second, to counteract too great a loss of information in the translation from numerical value to graphical symbol, as many discriminable symbol elements as possible should be used. Third, in order to keep the presentation of large data sets surveyable, the largest symbol to be used must be restricted in absolute size. The smaller the symbols, the more "data" can be presented in a graphical matrix of a given size. A fourth demand concerns the adequacy of judgement of the symbol sizes. It must be possible to correctly estimate the intended value or information of all separate elements that are used in the matrix. With these requirements in mind, the theory of Bertin can be studied in the light of cartographical and psychophysical research on the use of size of graphical symbols as an alternative method in the presentation of numerical information.

6.2.1. Detectability of differences in size

First of all, elements in a graphical matrix must be discriminable. When the original numerical values are translated into graphical point symbols, all "important" differences between these numerical values must also be discriminable in their graphical counterparts. In addition, the sizes of all graphical symbols that represent different categories or values must be identifiable.

In the cartographical literature there are two important methods of presenting and comparing quantitative data in a graphical format.

The first method is that of a direct conversion from basic quantitative value into the size of a graphical symbol, revealing a monotonous relation between authentic quantitative value and symbol size. Every change in numerical value, however small, results in a change in size of the symbol representing it. This is called the graduated method.

The second method is that of range-graded symbols. The total range of values is first divided into a number of categories, followed by attaching a symbol of a fixed size to every category and thus also to every value within a certain category. Both methods, as Meihoefer (1973) noted, have some advantages and disadvantages, the most important disadvantage of the use of graduated symbols being that slight increments in circle size make them difficult to distinguish. It is therefore probably even more difficult to attach correct numerical values to these symbols. Although there is a loss of information when the range-graded method is applied, this seems to be the more useful and effective one.

6.2.2. Number and maximum size of different elements

The demands governing a large number of different elements and restricted absolute size of the largest symbol are interrelated and are therefore discussed at the same time. Comparing a number of cartographical studies, it can be seen that although there is some degree of variation in the range of stimulus sizes used, the smallest circles generally have diameters of 2 to 3 millimetres. The size of the largest stimuli used is more subject to variation, with ratios of max. area/ min. area ranging from less than ten to several hundred (see Chang, 1980). Projecting these larger symbols in the cells of a matrix would require matrix cells with a side length of several centimetres, a very inefficient size for the presentation of large data sets. As regards the useful size of symbols, the cartographical studies don't offer a workable proposal. Bertin (1981), on the other hand, gives a more suitable maximum symbol size as a starting point. From a very practical standpoint, the size of circles on prefabricated plates, he proposes to restrict the size of the largest circle to one with a diameter of 5.0 millimetres. In chapter 5 of this thesis some experiments were discussed in which circles ranging in size from 0.2 mm^2 to 19.6 mm^2 (diameters ranging from 0.5 to 5.0 mm.) were tested for discriminability (In some of the experiments of Chapter 5, stimuli were presented on a computer screen. The actual sizes of these stimuli were much greater in centimetres but not in visual angle). Although the ratio between two discernible circles was related to their absolute size, results showed that a ratio in area of 1.12 was sufficient to keep the error rate of discrimination as low as about 10%. This would mean that with the largest circle having a diameter of only 5.0 mm. and the smallest stimulus having a diameter of say 1.0 mm., at least 15 circles could be differentiated ($1.12^{14} \approx 5.0$).

6.2.3. Judgement of size of individual symbol elements

The demand as to the possibility of judging the size or value of individual graphical symbols used in a matrix to represent numerical values is the most extensively studied of all the above-mentioned demands. It also is the most important and intricate one. Whereas testing symbols as to discriminability only requires a methodologically simple experiment, selecting a number of circles whose judged size equals its intended size requires a methodologically more intricate investigation. In discrimination tasks, differences in size between stimuli can always be made greater than necessary in order to play safe, but in the estimation of size, perceived differences need to be in accordance with actual differences. We will subsequently discuss some of the major findings regarding the estimation of size of various graphical symbols within areas of psychophysics, cartography and exploratory data analysis.

Psychophysics

In his famous 1957 article, Stevens made a distinction between two classes of perceptual continua (1957, p.154):

> Continua having to do with *how much* belong to what we have called Class I, or prothetic; continua having to do with *what kind* and *where* (position) belong to Class II, or metathetic. Class I seems to include, among other things, those continua on which discrimination is mediated by an additive or prothetic process at the physiological level. An example is loudness, where we progress along the continuum by adding excitation to excitation. Class II includes continua on which discrimination is mediated by a physiological process that is substitutive, or metathetic. An example is pitch, where we progress along the continuum by substituting excitation for excitation, i.e., by changing the locus of excitation.

The perceptual continua of length and area belong to the first group, Class I. In a large number of experiments that were performed or cited by Stevens, he showed that, for the Class-I continua, the relation between a subjective sensation (S) and the corresponding stimulus magnitude (I) could be described by a power function,

$$S = c \cdot I^n \qquad [6.1]$$

(where S is perceived sensation, I the actual stimulus magnitude or intensity and c the constant).

This power function can be more easily expressed in terms of logarithms. Using logarithms, this function can be represented as

$$\log S = \log c + n \log I \qquad [6.2]$$

(This linear equation is of the form $y = b + a \cdot x$)

An exponent n smaller than 1.0 would mean that, in a comparison between two stimuli, the perceived difference in magnitude of these stimuli will be underestimated.

Stevens distinguishes four methods for the construction of ratio scales of subjective magnitude:

- ratio estimation (the experimenter presents two or more stimuli and asks the subject to name the ratio between them);
- ratio production (the subject is allowed to adjust a stimulus to produce a prescribed ratio to a reference or the subject must answer whether two stimuli meet a predescribed ratio);
- magnitude estimation (subjects must assign numbers to a series of stimuli under the instruction to make the numbers proportional to the apparent magnitudes of the sensations produced);

- magnitude production (subjects must adjust stimuli to produce proportionate subjective values).

Stevens already noted a dependence between the magnitude of the power in the power law and the method used in the experiment. In addition, the value of the exponent was dependent on the perceptual continuum that was measured. Exponents ranged from 0.3 for loudness to 2.0 for visual flash rate. The exponent for visual length that is mentioned by Stevens is 1.1, whereas the exponent for visual area varied from 0.9 to 1.15. Although these values seem to indicate that the size of stimuli can be estimated fairly well, other experiments of Stevens (1975) and Stevens & Guirao (1963) revealed exponents that were considerably less than unity.

In a study by Ekman & Junge (1961) the subjective length of lines, area of squares and volume of solids and drawn cubes were measured. The exponents of the power functions for these four conditions were respectively: 1.11, 0.91, 1.01, and 0.79. A second series of experiments within the same study investigated the influence of the stimulus range on the exponent. Squares and circles were used, with areas ranging from 1:2.1 to 1:9.5. In the case of squares, the power function exponents of the series appeared to increase somewhat with range (exponents ranged from 0.98 to 1.06), exponents of circles did not appear to vary with stimulus range. In all the experiments a ratio estimation of the presented stimuli was used. In some of the experiments the stimuli were presented pairwise, in others a larger number of stimuli were drawn on a chart. The symbol that was marked had to be used as a reference.

Whereas some studies confirmed the results of Ekman et al., also yielding high exponents for size judgements (e.g. 0.96, Sjöberg, 1971; and 0.99, Baird, 1965), others found considerable underestimation of the actual area of circles (as e.g 0.75, Marks & Cain, 1972; 0.70 - 0.76, Mashour & Hosman, 1968; and 0.76, Teghtsoonian, 1965).

MacMillan et al. (1974) proposed that the differences in the exponents that were found in different studies were mainly due to the presence or absence of a reference. Their own experimental results showed larger exponents when a reference figure was presented during judgement of the size of the stimuli. The effect seemed to occur in both magnitude estimation and magnitude production and was independent of the relative size of the reference (compared to the size of the estimated stimuli).

In a study comparing two kinds of subject instruction, Teghtsoonian (1965) found that an instruction emphasizing the *areal* property of circles gave more accurate judgements of size than an instruction emphasizing apparent *size*. The slopes of the power functions were respectively 1.03 and 0.76. Thus, when subjects were asked to judge the apparent *size* of two-dimensional figures, the exponents of the resulting power functions were appreciably less than unity.

How repeatable Steven's power law exponent is for individual subjects was investigated by Teghtsoonian and Teghtsoonian (1971). Only at very short time intervals between sessions, were correlation coefficients (Pearson r) between individual power functions found to be significant. The authors concluded that the low correlation values indicated that at least 90% of the variance in individual exponents should be attributed to chance factors. The real individual differences in exponents that were found could not be regarded as enduring individual characteristics.

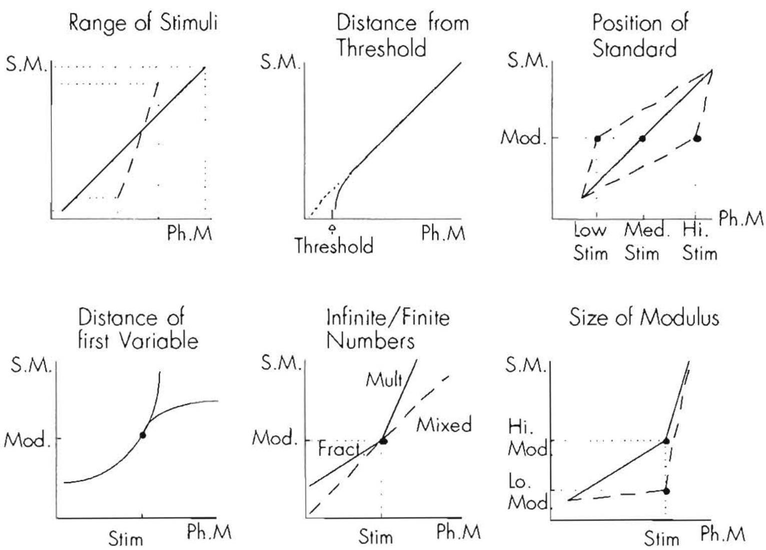


Figure 6.1 Different models showing the influence of independent variables upon the function obtained by magnitude estimation. (S.M: Subjective magnitude, Ph.M.: Physical magnitude, Mod.:Modulus, Stim.: Reference Stimulus (after Poulton, 1968)

Poulton (1968) did not concentrate on the influence of one single aspect in the estimation of symbol size, but systematically changed a number of aspects that could possibly have an influence on the exponent in the power law. He discussed 6 models that describe the relation between values of various independent variables and magnitude estimation (see Figure 6.1).

1. Range of experimental stimuli. A shorter range of stimuli produces a steeper function than a longer range of stimuli.

2. Distance from the threshold. The slope of the function appears to be steeper near the threshold.
3. Position of the reference. When the physical magnitude of the reference is near the lower end of the range, variables smaller than the reference give steeper slopes than those larger than the reference. The opposite is found when the reference is near the upper end of the range; now the variables that are larger than the reference result in the steeper slopes. Poulton suggests selecting a reference in the middle of the range, such that the slope for the variables larger than the reference is the same as the slope for the variables smaller than the reference.
4. Distance of the first variable from the reference. The steepness of the function is dependent not only on the distance from the reference, whether the variable stimulus is larger or smaller than the reference, but also on the past experience of the observer.
5. Infinite versus finite sets of stimuli. Using only multiple stimuli (stimuli larger than the reference) and an infinite set of numbers available to the observer, gives steeper slopes than fractional estimates (stimuli smaller than the reference) which are limited at the end of the set by zero. When a combination of both multiple and fractional stimuli are used, an intermediate slope is found, less steep than the multiple estimate function and steeper than the fractional estimate function.
6. Size of the modulus (number given to the reference). Increasing the modulus increases the set of numbers available for fractional estimates, and thus increases the slope. Conversely, it reduces the set of numbers that are commonly used for multiple estimates, and thus decreases the slope.

Concluding this enumeration of studies, it is found that the estimation of size of visually presented stimuli can be adequately described by Steven's power law. The exponent of the specific functions, however, is largely dependent on a number of methodological characteristics of the experiment. In addition there is some evidence that the exponent is a decreasing function of the number of spatial dimensions of the stimulus.

Cartography

Whereas psychophysical studies regarding the estimation of size of visually presented stimuli were mainly interested in investigation, description and explanation of psychological and perceptual functions, the cartographical object of studies on this topic is that of the adequacy and accuracy of judgement. Running parallel to psychophysical studies, in the area of cartography the first experiments which give evidence of underestimation for a number of point symbols which vary in areal size, date back to the late 1950s.

When graduated circles were presented on cartographical maps, Flannery (1956) found a power function exponent of 0.87 in the estimation of differences (ratio estimation) between these circles. These findings were apparently so convincing to Robinson and Sale that they decided to insert a table in their (standard) book on cartography (Elements of Cartography, 1968) to simplify the conversion of the actually intended size of circles to their apparent estimated size. This conversion rule was directly based on Flannery's power function exponent. Others were obviously less convinced by the results of Flannery's study, as there has been no widespread acceptance of this system.

In a number of experiments, performed with the intention of not only retesting earlier found exponents, but also comparing the effectiveness of circles to other point symbols, Flannery (1971) again showed a consistent underestimation of circle symbols on thematic maps. Exponents that were found in these experiments, 0.86 for circles and 0.82 for wedges, showed a close resemblance to the results of earlier studies. Wedges were considered to be less effective than circles because the estimation of their size was not as consistent, due to a greater variance. Tests on the apparent size of bars showed no underestimation, the percentage of overratings being about equal to the percentage of underestimations. In these studies the method of ratio estimation was again applied.

In a study by Crawford (1971), the experiments of Flannery were replicated except for the grey tone of the symbols. Whereas stimuli in the Flannery experiment were solid black circles, Crawford varied the percentages of grey tone of the symbols between conditions. Results indicated that all grey tones (these grey tones varied from 30% to 60% black) transmitted visual signals that were not statistically different from the visual signal transmitted by black graduated circles.

One of the first experiments that started to show some attention to methodological aspects of the experimental design was performed by Meihoefer (1969, 1973). In his experiments he studied the effect of a systematic change in the number and size of reference stimuli. In the first experiment it was shown that underestimation of the size of a circle increased as the difference in size from the reference circle (the smallest circle acted as a reference in this experiment) became greater. A following experiment, using a more extended standard containing the two extreme values, showed an improvement in the estimation of circle size compared to the first experiment. In a third experiment, in which a map legend was used that contained all circles that appeared on the map, the error in circle size estimation was either zero or clustered around zero. It seems obvious, therefore, that the provision of more information by using more circle sizes in the legend results in less ambiguity or a smaller error in estimation. These findings are confirmed in other studies (e.g. Dobson, 1974; Cox, 1976). When using a reference symbol, Cox claimed that two counterproductive effects of the reference symbol are operative in size judgement: a contrast effect resulting in a displacement of judgement away from the reference (overestimation of circles larger than the reference and underestimation of circles smaller than the reference) and an assimilation effect tending to a displacement of judgements in the direction of the reference (underestimation of circles with a small reference and overestimation with a large one). Results indeed showed a greater underestimation when a small reference was used, a slight overestimation when a large reference was used and fairly accurate estimates with a middle-sized one or with all three reference circles. Again, the results are in accordance with one of the models of Poulton (1968). The standard deviation in the condition using three legend circles was significantly smaller than that in the condition using the middle-sized reference. Concluding the role of the reference, it can be confirmed that an extensive reference containing the extreme circle sizes, as well as some intermediate circles, will give both the most accurate overall judgements and the least variance in size judgements.

Comparison of a number of studies that used the method of magnitude estimation with a prescribed modulus (what is the size of a test circle knowing that a reference circle has a size of 100?) with studies using direct-ratio estimation (how much smaller/larger than the reference circle is the test circle?) by Chang (1980) showed that magnitude estimation is more difficult than ratio estimation. With ratio estimation the power function exponents were closer to a value of 1.0, which corresponds to a correct judgement of size. Other factors mentioned by Chang, that are of influence on the exponent are given below.

- Instruction given to the subject. It seemed that the subject's awareness of the areal property of circles could result in better estimates and thus increase the exponent.

- Role of the reference. Chang notes that an extension of the number of keys in the reference changes the task of the subject from magnitude estimation to a twofold task. In the first part the stimuli are grouped according to categories formed by the standard keys. In the second part magnitude estimates are made.
- Stimulus range. An increase in the stimulus range can result in a decrease in the exponent.
- Sequential effect. Estimation of a given stimulus can be influenced by the magnitude of the previous stimulus or stimuli that are estimated (Cross, 1973). This effect is found to be strikingly effective when a number of stimuli have to be estimated sequentially and the reference is only shown at the beginning.

Other factors possibly influencing size judgement that are not further discussed here are, for instance the effect of surrounding circles on the size estimation of a target circle, the so-called Ebbinghaus illusion (discussed for example in Massaro & Anderson, 1971), effect of experience or the role of training and feedback (Olson, 1975b) and of repeatability of results or the session-to-session correlations of size judgement on subjects (Teghtsoonian & Teghtsoonian, 1971).

Exploratory Data Analysis

Cleveland, Harris and McGill (1982, 1983) used circles of different sizes on statistical maps. Subjects were told that these circles represented the average daily long-distance telephone charges of different companies. The size of one marked circle represented $100 and subjects were asked to estimate the charges of a number of circles on the map (magnitude estimation). Although there was a large variability across subjects, results showed fitted power functions for individual subjects that were mostly close to 1.0 (the median exponent was 0.96, upper and lower quartile 1.00 and 0.84 and the range of exponents 0.58 - 1.27). Furthermore, no differences were found between a group of scientifically trained subjects and high-school students, between a map-like (grid ticks, labels, scale, border) and no map-like condition or between the instruction to estimate dollars versus the instruction to estimate the area of the stimuli.

Towards an experimental design

Reviewing the demands made on characteristics of symbol elements a number of pertinent questions can now be examined in the present study.

The first concerns the question as to the possibility of giving correct estimates of the size of small symbols presented in a matrix. This is a variation on most psychophysical and cartographical studies in the size and configuration of the stimuli. In our experiment, relatively small symbols (0.5 -5.0 mm. diameter) will be used in a strict matrix format. In accordance with a number of studies which use an extended legend, a legend of 5 symbols, including the two extremes, is used in the present study. It is to be expected that overall estimation errors will be quite small as, in addition to the use of a legend, differences in size of the range-graded symbols used will be made large enough to be correctly detectable (see Chapter 5).

When studying the raw data and median estimates of a number of studies (e.g. Meihoefer 1973, Crawford, 1971, Cleveland et al., 1983) we noted a tendency in subjects to use whole numbers or even multiples of five or ten in their estimation of stimulus size. In the study by Crawford (1971) for example, actual size differences of 3.45 and 3.32 resulted in a median of estimates of 3.00, and differences of 5.97 and 6.18 in a median estimate of 5.00. This would mean that slight variations in the exponent can be obtained by careful selection of the size of the stimuli used. This finding is also associated with Chang's (1980) observation that subjects tend to group stimuli according to categories before making actual magnitude estimates. It would be wise to take account of this tendency when selecting various sizes of point symbols for use in a matrix.

Another question concerns a variable not hitherto mentioned— the grid line. The question is whether the presence or absence of grid lines in a matrix has a differential effect on the precision of size estimations. The possible influence of the use of grid lines is an interesting one, because this method is not accessible in a map-like context. In a matrix, on the other hand, it is possible to use grid lines next to a legend or standard, as a second kind of reference when estimating the size of a symbol. Not only can the symbols in the matrix be compared with each other or with the key values in the legend, but the distance from the outer edge of a stimulus symbol to the grid line enclosing its matrix cell or the amount of white space surrounding the symbol in the cell can also be used in gathering information on the correct size of the stimulus. This added source of information could also be very helpful when comparing different cells or symbols in the matrix. It is therefore expected that the presentation of grid lines will have a positive effect on the accuracy of estimates of graphical symbols in the matrix. This expectation is contrary to the results of an experiment by

Dobson (1980). In a matching task in which the size of a tachistoscopically presented fixation symbol had to be compared with the size of two identical targets presented at equal distances to the left and the right of the target, Dobson found a detrimental effect of grid lines. Accuracy of matching decreased as a function of the graphical information content (cell, line or map noise) of a fixated scene and as a function of the distance separating the targets and the fixation symbol. With tachistoscopically presented stimuli, the addition of extra information apparently acts as a disturbing factor. Because there is no time limit in the experiment discussed below, a positive effect from the amount of graphical information is nonetheless expected.

A third question concerns the shape of the symbols. Is there a difference in the estimation error between symbols of different shapes? It is quite possible that the size of some symbol shapes can be more accurately estimated than that of other shapes. A number of studies showed that size judgements of "unidimensional" shapes, such as the length of a line or the height of a bar are more accurate than judgements of "two-dimensional" shapes, such as circles or squares. Unidimensional estimates generally result in a power exponent closer to 1.0 compared to two-dimensional estimates. In other words, the power function exponent is a decreasing function of the number of spatial dimensions (Teghtsoonian, 1965). Studies comparing two and three-dimensional shapes, such as cubes and spheres showed an even greater estimation error for the three-dimensional shapes (Ekman & Junge, 1961; Ekman, Lindman & William-Olsson, 1961). Since bars, as well as circles and squares will be used in this study, it is expected that the estimation of bars will show a slight error rate compared to those of circles and squares.

6.3 Experiment 1

Stimulus material

Three sets of symbol shapes were used; circles, squares and bars, each set consisting of 16 elements of different size. All elements were solid black. Thirty elements of the same shape were presented in a 6 · 6 matrix (or 36 in a 6 · 7 matrix in order to present each symbol the same number of times overall) with cell sizes of 5 mm square. One cell was left empty in each row and column of the 6·6 matrix. Sizes of the symbols were randomly assigned to the cells of a matrix, with the restriction that no stimulus size occurred more than three times within a certain matrix. All symbol sizes were presented 6 times. Furthermore, two conditions were used. In condition 1, the symbols were presented in a 3 · 3 cm. square, with horizontal and vertical grid lines projected between the 0.5 · 0.5-cm- square cells. In condition 2, no grid lines were drawn, leaving only the contour lines of the square. Examples of the two conditions as presented in the experiment are given in Figure 6.2. With all subjects getting the same 18 matrices in two sessions in a counterbalanced order, the design of the experiment can be defined as a crossed, counterbalanced, repeated-measurements design.

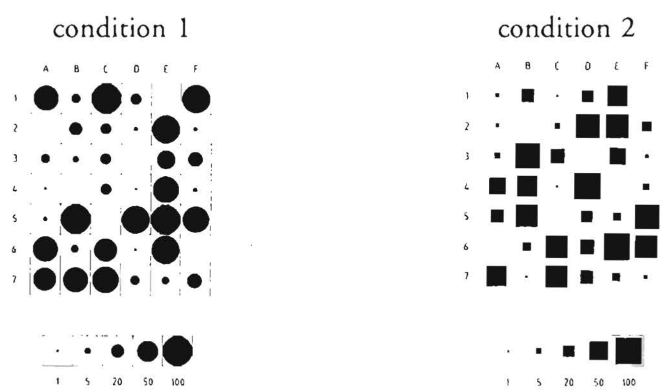


Figure 6.2 Examples of matrices with legend key as presented in the experiment

About four inches below the matrices a legend was given, consisting of 5 symbols covering the complete range (including the minimum and maximum sizes). The legend was not projected directly beneath the matrix to make an eye movement necessary in the comparison of legend symbols and matrix symbols, so that the matrix symbols nearest the legend are not advantaged compared to the matrix symbols further away from the legend. Dobson (1977) notes that, "while fixating a map at a reading distance of 12 inches, a subject can clearly see an area approximately 1.47 inches in diameter. In order to view a new area outside of this foveal cone of clear vision (of about 7 degrees of visual arc) the eye must move to that location."

The area of the smallest circle and square and the height of the smallest bar were given an arbitrary value (modulus) of 1. All other symbols had a value directly proportional to the size of these smallest symbols. Hence the largest bar, square and circle with values of 100 were exactly 100 times as large as the smallest ones. The symbols were respectively 2.5, 5, 7, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90 and 100 times as large as the smallest symbol.

This means that the ratio of subsequent elements in the series ranged from 2.5 (2.5:1) to 1.11 (100:90) These values are in accordance with correct discrimination of stimuli of 90% or more (see Chapter 5).

Procedure

The experiment was a self-paced paper-and-pencil test. In each of the two conditions (grid and no grid) a booklet containing nine cards was offered, three cards for each of the symbols (dots, squares, bars). The order of the cards was counterbalanced. Subjects were asked to study the matrices sequentially and estimate the sizes of the symbols projected in the cells. Estimations were entered in the cells of a table on a response form. Except for the empty cells of this response table, it was identical to the test matrix.

Subjects

Twenty-one subjects from different faculties of the Eindhoven University of Technology participated in this study. For every subject the experiment consisted of two sessions, with an interval of at least one day between sessions. The conditions or sessions were counterbalanced across subjects. Within conditions, all subjects received the same matrices in a counterbalanced order. Subjects were advised to take short breaks after each matrix, but not to linger too long in their judgements of the individual stimuli. Each condition took 45 to 65 minutes. Subjects were paid for participation.

Results and Discussion

1. Power law exponents

In the first part of the analysis log-log functions between actual and estimated symbol sizes were calculated. The exponents of individual power-function exponents were all found to be very close to the optimum of 1.00. In Table 6.1 some examples of the exponents for various subjects are given, together with the maximum and minimum exponents, the mean of exponents and standard deviation.

Table 6.1. Exponents of power functions for 5 of the 22 subjects (columns 1 to 5 inclusive), the range of exponents for all subjects (maximum and minimum), their mean and the standard deviation of the mean. These values are shown for each of the experimental conditions.

	Subjects								
condition	1	2	3	4	5	Min	Max	Mean	S.D.
dots grid	.99	1.00	.97	1.02	.95	.92	1.02	.99	.025
dots no grid	1.01	.96	1.01	.95	.91	.91	1.02	.99	.031
squares grid	.99	.95	1.01	.96	1.00	.94	1.04	.98	.030
squares no grid	.95	.99	1.00	1.00	.97	.88	1.03	.98	.038
bars grid	.93	.87	.90	.86	.93	.84	.97	.92	.033
bars no grid	1.02	1.00	.99	.99	1.01	.91	1.04	1.00	.025

In Figure 6.3, power functions are given for each of the three symbol shapes. Functions are obtained by including estimates of both grid and no-grid conditions. Exponents are very close to the value of 1.0, and the two symbols, representing the grid and no-grid conditions, that are plotted at each of the stimulus scale values nearly all coincide, indicating only a slight difference between these two conditions. The correlation coefficients between the log values of the actual size and the log values of all individual estimates for the dots, squares and bars were respectively 0.98, 0.98 and 0.97, which means that almost the complete variance of the estimates is associated with the actual symbol size. The standard error of estimates were, respectively 0.08, 0.09 and 0.08 (all $p<0.001$).

The presence of an extended legend containing both maximum and minimum, as well as some intermediate stimulus sizes appears to be sufficient to avoid a general tendency to underestimation of differences between stimuli.

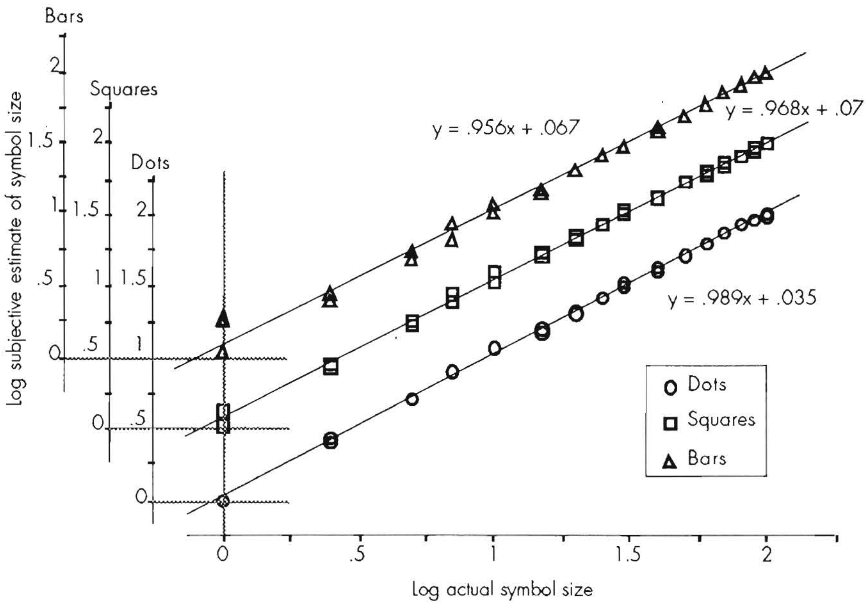


Figure 6.3 Log values of the estimated size of the for dots, squares and bars plotted against the corresponding logs of actual values. At each stimulus-scale value the average subjective values of grid and no-grid conditions are plotted. Each of the three y-axes belongs to one of the symbol shapes.

2. Estimation error

Estimation errors as percentages of the real areal symbol size were calculated as the crucial dependent variable in this second part of the analyses. Estimation error can be defined as the ratio of the subjectively judged size (X_s) minus the objective real size (X_o) of a stimulus (X_s - X_o) and the objective real size. In order to get percentages, this ratio has to be multiplied by 100.

$$\text{Estimation error, } x = \frac{X_s - X_o}{X_o} \cdot 100 \qquad [6.3]$$

A positive estimation error therefore means an overestimation and a negative value an underestimation of the real stimulus size.

In addition to these relative estimation errors, their absolute values (absolute estimation error) were also used in several analyses because the direction of the estimation error sometimes acted as a kind of confounding variable. Averaging large overestimations and underestimations can result in a mean correct estimation, thus with an mean error of nil.

Absolute estimation error, $|x| = \frac{|X_s - X_o|}{X_o} \cdot 100$ [6.4]

In Table 6.2, mean estimation errors, mean absolute estimation errors and their respective standard deviations are shown for each of the experimental conditions.

Table 6.2. Means and standard deviations of the estimation error, x, and absolute estimation error, |x|, for each of the experimental conditions.

	Statistics			
condition	Mean of x	St. Dev of x	Mean of |x|	St. Dev of |x|
dots grid	8.2	17.5	11.0	16.3
dots no grid	4.9	15.5	9.9	14.1
squares grid	9.8	16.6	12.7	15.2
squares no grid	5.1	17.4	11.8	15.3
bars grid	3.1	15.6	8.0	12.6
bars no grid	-0.7	12.0	6.6	10.1

Grid versus no-grid condition

For the mean estimation error, the differences between grid and no-grid were significant for all three symbol shapes (all $p<.001$). For the absolute mean estimation error for only the bar symbols, the difference between these conditions was significant (bars, $p<.01$; dots and squares, $p>.02$).

In the left-hand part of Figure 6.4, grid and no-grid condition are compared for the three symbol shapes separately. In this figure it can be seen that the mean overestimation is greater for the grid condition with all symbol shapes. These findings seem to be contrary to the hypothesis that the presence of grid lines has a positive effect on the size of the estimation error. In the right-hand part, the mean absolute estimation error is shown for all conditions. Differences between grid and no-grid conditions are slighter.

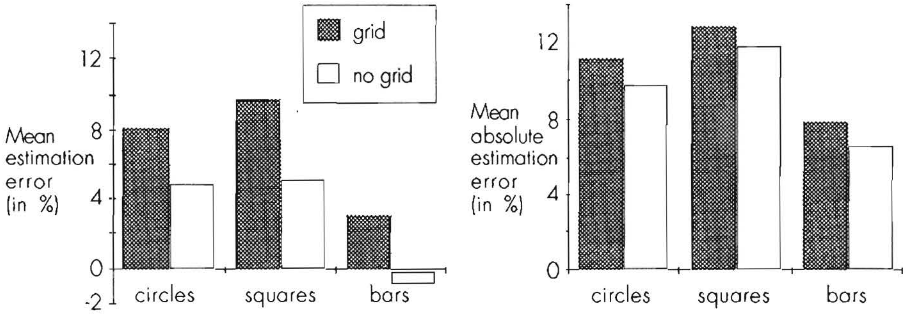


Figure 6.4 Average estimation errors (left-hand part) and average absolute estimation errors for the grid and no-grid conditions for each of the three symbol shapes. Estimation error and absolute estimation error are expressed as percentages of the real size of the symbols.

Differences between symbols

All pairwise comparisons of symbols revealed significant differences (all p<.01) except for the difference in mean estimation error between dots and squares (p>.1). The most striking differences, however, are those between bars on the one hand and circles and squares on the other. The mean *absolute estimation error* of 6.6 for the bar, no-grid condition, shows that, on average, estimations of subjects deviated 6.6% from the real areal size. Since the mean *estimation error* for the same condition is -0.7, there were apparently more underestimations than overestimations for this condition. For all other conditions the number of overestimations exceeded the number of underestimations.

Another noteworthy result is elucidated in Table 6.3. In this table the frequency of estimation errors is distributed across the different stimulus sizes. The sizes of the errors are almost all round values, as were the real sizes of the different symbols presented. From this table it can be concluded that the subjects in the experiment had a disposition to using round numbers (multiples of 5 or 10) when estimating the areal sizes of graphical symbols in the matrix. Stimuli with an actual relative size of 7 were, for example estimated correctly in only 15% of the presentations. In not less than 38% the estimated size of these stimuli was 10 and in 22% the estimated size was 5. These findings are in accordance with our hypothesis and the previously mentioned observation of Chang (1980). It should be noted that, apart from the sizes of the symbols in the legend, subjects were not informed about the different symbol sizes used in the experiment, nor did they know whether graduated or range-graded symbols were used.

Table 6.3 Distribution of estimation errors in percentages, averaged over dots, squares and bars and over grid and no-grid conditions

Estim. error	Actual stimulus size 1	2.5	5	7	10	15	20	25	30	40	50	60	70	80	90	100
-20												1	1			
-15										1					1	
-10							1		2	4	3	12	5	5	6	4
-5 →					2	12	6	16	16	11	5	3	4	7	5	5
-3					1	3	1									1
-2.5			1		1	1										2
-2			2	22	1			1								3
-1	5			1												3
-0.5		6														
0 →	75	32	84	15	61	53	72	45	52	62	78	57	52	56	65	83
0.5	1	46		8												
1	5		1	10		1										
1.5	13	5														
2			3	1	3	2	1									
2.5		11	1													
3			1	38	1	2										
5 →			7		25	25	14	32	12	6	2	6	14	8	12	
8				3												
10 →			1		3	1	4	2	14	15	10	14	16	19	9	
15								2				3	3	1		
20									2	1	2	3	3	3		
25								1			1					
30												1				

In Figure 6.5 a distribution of estimation errors is given for each of the different symbol sizes (sizes of the symbols varied from 1 to 100 inclusive). Errors are shown for grid (gray bars in the figure) and no-grid (white bars) conditions. The estimation errors are averaged over dots, squares and bars. The curve in this figure displays the differences between the values of grid and no-grid conditions. It can be seen that the largest errors (in percentages of the actual symbol size) were made with the smaller symbols, and notably in the grid condition. From size 20 upwards the size of the estimation error is quite stable. Differences between values of the grid and no-grid conditions are relatively slight near the legend values (encircled values on the x-axis) and increase as the difference in size between stimulus and nearest key value increases.

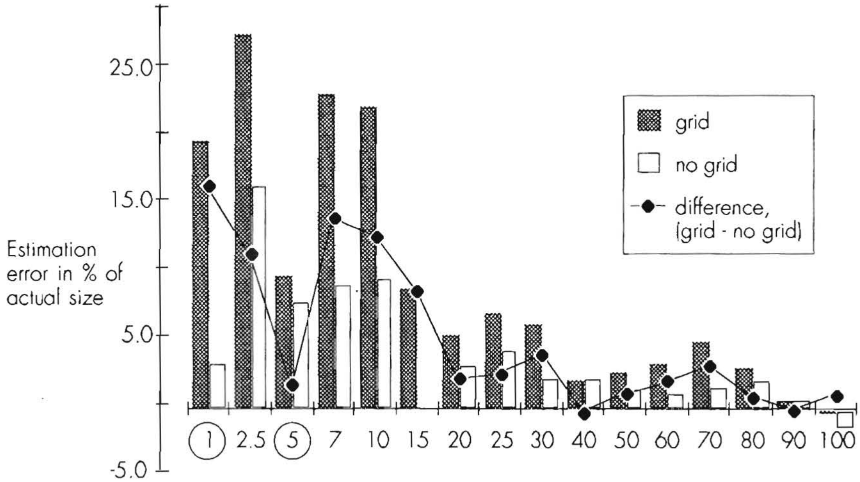


Figure 6.5 Estimation error for each of the symbol sizes on the x-axis, for both the grid (gray bars) and no-grid (white bars) conditions. Values are obtained by averaging over the symbol shapes. The curve indicates the difference between the values of the grid and no-grid conditions. Encircled numbers on the x-axis indicate the values that were included in the legend.

6.4 Conclusion

Subjective estimates of the size of symbols can be very well described by a power function. All exponents of the curves for the dots, squares and bars are close to an optimum of 1.00 and standard deviations of residuals (portion of estimated values not associated with the actual values) can be neglected.

As the exponents of the power functions were all close to a value of 1.00, it is very improbable that Bertin's value of 1.12 is deduced from power function exponents.

If estimation errors are expressed as percentages of the actual stimulus size, bar symbols provide the most accurate estimations. The size of dots and squares is generally slightly overestimated. Linear measures (bars) are probably easier to estimate than areal measures (dots and squares).

The presentation of an extended reference key is recommended, as the size of the estimation errors decreases near the values of the reference keys. When all sizes of the symbols used are included in the reference key, the task of subjects changes from an actual estimation of the size of a stimulus by interpolation to one of comparing and correctly recognising target stimuli and legend keys.

Subjects show an obvious tendency to use round numbers, for instance multiples of 5 or 10, in their estimations of the size of graphical symbols. This disposition seems to be less dependent on the actual size of the presented symbols. The tendency to use round numbers links up with and can be satisfied by application of the range-graded method.

The use of grid lines has to be discouraged, as it results in larger estimation errors of the actual size of stimuli compared to the condition in which these grid lines are absent.

Chapter 7. Structure and pattern recognition within a graphical matrix

7.1 Abstract

The functioning of the graphical matrix at the intermediate (groups or clusters of circle symbols) and overall (complete matrix) level of information processing is investigated in this chapter.
First an experiment tested whether subjectively perceived clusters in the matrix could be predicted by what was originally a cartographic model based on three characteristics of the map and the symbols displayed on it. The three model variables are based on Gestalt principles. Results showed that all three variables of the model were related to the perception of clusters although this relation was not strong enough to correctly predict all matrix elements as either clustered or unclustered. In particular, many of the perceived unclustered elements were wrongly estimated as they were predicted to be part of a cluster. As most of these wrongly predicted groups were either notably smaller than the perceived clusters or of a shape that can be regarded as irregular, variables related to the number of symbols in a group and to the continuity in shape of the clusters, respectively, might prove to be useful. At the overall level, two experiments were performed on the subjective ordering and rating of orderliness of matrices. Results revealed that there is notable agreement between subjects and between sessions regarding this orderliness. In addition, subjective ratings were accurately predicted by a simple model that is based on geometric distances (differences) between graphical symbols in adjacent matrix cells.

7.2 Introduction

We have demonstrated in the preceding chapter that subjects are able to give fairly correct estimates of the size of graphical symbols when these symbols are presented in a matrix and are accompanied by a number of keys in an appended legend. A mere correct identification of individual symbol elements, however, does not seem to provide too much of a surplus value of the graphical matrix compared with a numerical matrix. Ordering differently sized symbols takes less time but depends on bigger differences between the considered elements compared to the ordering of numerical values. Sizes of graphical elements can be estimated fairly easily and accurately, but here again, the numerical counterparts are far more accurate.

We could conclude that, as long as precise ascertainment and discrimination of individual elements is essential, numerical presentation indubitably has some surplus value compared with a graphical presentation. When, however, the exact values of the items are less essential, when a large number of comparisons has to be made, or when the time required to discriminate elements is of importance, then the graphical matrix seems to provide the better option. The scale clearly turns in favour of the last-mentioned alternative when the processing of information at higher levels is drawn into the comparison. The relative efficacy of the graphical matrix becomes more obvious when we consider the information within the matrix at the intermediate and overall levels.

We have already stated in Chapter 1 that one of the requirements of the graphical matrix is the possibility of a fast, correct interpretation of the overall picture or the total information that is displayed in the matrix.

Bertin (1983) claims that information presented in a graphical matrix can be processed at various levels. Besides an analysis at the elementary level (see also Chapter 4) interest can also be directed at (the hypothesis stated a priori) or attracted to (the a posteriori or visual method) a particular matrix subset. When elements of various sizes are scattered all over a matrix, a contiguous group of symbols of similar size is not only bound to draw the attention of the observer, but also brings a kind of order into the chaotic whole. At the level of the overall image, the complete matrix can be evaluated by rating the general distribution of individual elements.

A more practical description of the presence of coherent groups of similar-sized graphical symbols is a situation in which some juxtaposed objects (in the columns of the table) contain analogous profiles of scores on the characteristics (in the rows). The same could apply to the characteristics component; when the strongly related characteristics (showing the same profiles along the object's component) are juxtaposed within the matrix, this will result in coherent groups of similar symbols.

In this chapter, the possibilities and strength of the graphical matrix at the intermediate and overall levels of information processing will be investigated. At the intermediate level, visual recognition of clusters or patterns of similar symbols will be further explored. We will check whether certain psychophysical properties of the symbols in the matrix efficaciously explain the patterns perceived there. At the overall level we will investigate whether the visually estimated degree of overall structure of a matrix is either a more general concept agreed upon by a majority of judges, a more ambiguous notion varying uncertainly between various judges, or merely a changing nuance of opinion in the mind of one and the same critic.

7.3 Recognition of clusters

When graphical symbols are presented in a matrix, some of them probably seem to belong together because of some visual properties of the symbols themselves and of their immediate environment on the map. When more or less similar symbols are concentrated in a particular part of the matrix, subjects tend to label this collection of symbols as a more or less coherent group, cluster or pattern. In this section we will look more closely into this phenomenon.

A study by Jenks (1975) inquired into several aspects of the visual interpretation and comparison of maps. Some of the problems investigated in this study and specifically of interest to our study concern

- the attributes of map symbolization and design that assist readers in perceiving spatial patterns and
- individual differences in the perception of clusters. Do map readers as a group see similar patterns on a map or is the regionalizing process highly individual?

In the first experiment of Jenks's study, 20 groups of experimentally controlled dyads (pairs of symbols) or triads (groups of 3 symbols) of circles were constructed. In both triads and dyads the circles were either presented tangentially or radially spaced (See Figure 7.1).

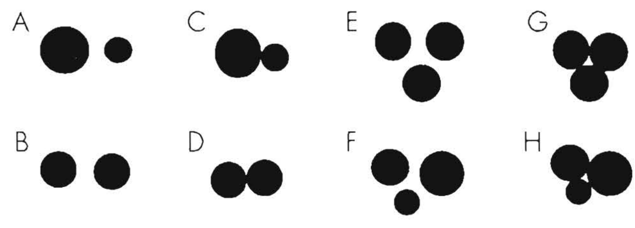


Figure 7.1 Dyads and triads created to give different visual impressions of clustering. A and B are examples of radially spaced dyads, C and D of tangential dyads, E and F are radially spaced triads and G and H tangentially spaced triads.

The total amount of black (a summation of the areas of the symbols) was the same in all triads; for the dyads it varied from one set to the other. Sets of circles were presented at the same time and their degree of clustering had to be rated on a five-point scale (ranging from 5 (very clustered) to 1 (very unclustered)). The resulting ordering of sets was more or less the same as in the following series, decreasing in cluster strength, that is tangent triads, tangent pairs, spaced triads and, lastly the set of spaced pairs. The results indicated that map readers did, in fact, see sets of circles in clusters, as all sets had median or mean values that were clearly greater than 1. An even more important result was that map readers were reasonably consistent in their visual ranking.

In a second experiment within the study, proportional circles were projected on cartographical maps. The total amount of black was approximately the same for all maps, whereas the total number of circles on each map and their respective sizes varied considerably.

In this experiment, respondents were directed to create clusters by drawing lines around a group of elements that all belonged to a region or "area of sameness" (p. 316, see also Figure 7.2). The percentage of subjects that selected the individual circles as being part of a cluster were calculated for each of the circles by examining and comparing isolines, each isoline representing the decision of an incremental 10 percent of the sample. Isolines are lines on a map joining places (here, graphic elements) of the same height (here, percentage clustered). Jenks concluded that, in accordance with the results of his first experiment, the group of subjects again saw distinct regions in the proportional circle maps and showed a high degree of agreement on the boundaries of the visual regions (examples of the isoline method are given in Figures 7.2 and 7.4). In order to further quantify these results, the relation between the perceived clustering and two physical characteristics of the maps and their symbols were investigated. The first hypothesis tested whether frequency of clustering was related to the potential surface (Stewart, 1947) of the symbols for any of the presented maps. The potential surface is a derived measure that involves the size of, and the distance between circles on the map. Correlations between the frequency with which stimuli were clustered and their potential surface value ranged from 0.62 to 0.92, with a mean of 0.81. The second variable measured the total amount of blackness in the immediate environment of each of the circles. The idea behind this hypothesis is that readers are unable to see details over the whole area of the map in a single (eye) fixation.

Lines around regions therefore have to be drawn in segments. Jenks proposes that these segments of boundaries fall within areas that can be seen with a high degree of clarity during a single fixation and that these boundaries are probably drawn between areas with significantly different degrees of blackness (Jenks, p. 318-319). The relation between the number of groupings of circles and the blackness of the fixated environment produced correlation coefficients ranging from 0.69 to 0.91, with a mean of 0.85. Combining the results of both measures, it was concluded that distance between circles, area of circles and the immediate environment around a given circle all play a part in verbal clustering. It is unfortunate that Jenks failed to examine the correlation between the two measures that he used. When we compared the two correlation coefficients that were found for each of the 6 different maps, it showed that their orderings were nearly identical. This could mean that both measures are very strongly interrelated and are, in fact measuring the same underlying concept.

Slocum and Gilmartin (1979) noted another restriction in the isoline method used by Jenks. This method is based on how often individual elements (circle symbols) are regarded as part of a cluster, but does not indicate the cluster or clusters to which the symbols belong (see Figure 7.2).

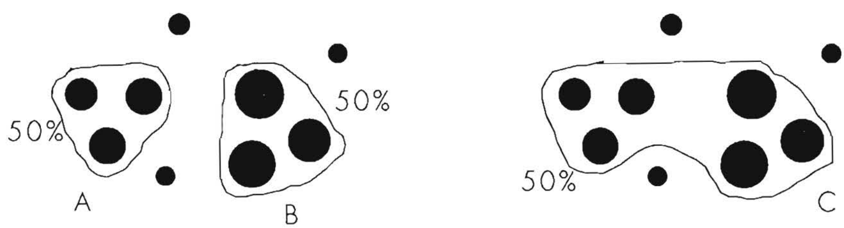


Figure 7.2 Within the group of elements presented on the left-hand side, two separate clusters, A and B, are encircled in 50% of the presentations of the stimulus set. If the isoline method were to be applied to display the clusters found in this set of elements (right-hand side), only one large cluster, C, would be encountered.

The isoline method is suitable for correct demarcation of clusters so long as the location of these groups is set wide apart on the map, with a number of circles clustered less often in between. Whenever two clusters are in the neighbourhood on a map, the isoline method will no longer recognize the two as being separate groups but will regard them as a single one. To get round this restriction, Slocum and Gilmartin proposed a new method for marking the intensity of a relation between elements on a map. This method is based on Thiessen polygons and geometrically related Delaunay triangles (Figure 7.3). In this approach, the analysis is no longer based on individual circles but rather on the linkage between two circles in a pair (see Figure 7.4).

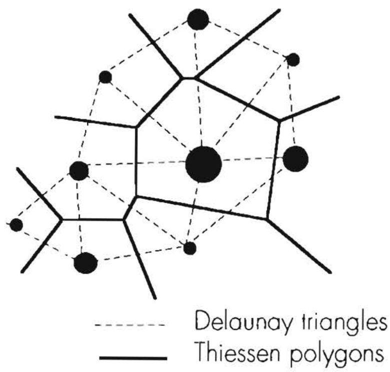


Figure 7.3 An example of the construction of Thiessen polygons and Delaunay triangles. With Thiessen polygons the geometrical plane is divided into polygons around the circle symbols which are situated at the centres of gravity. Every location within the resulting boxed polygons is closer to its own centre of gravity than to any other centre. Delaunay triangles are created by connecting the central points of all neighbouring Thiessen polygons.

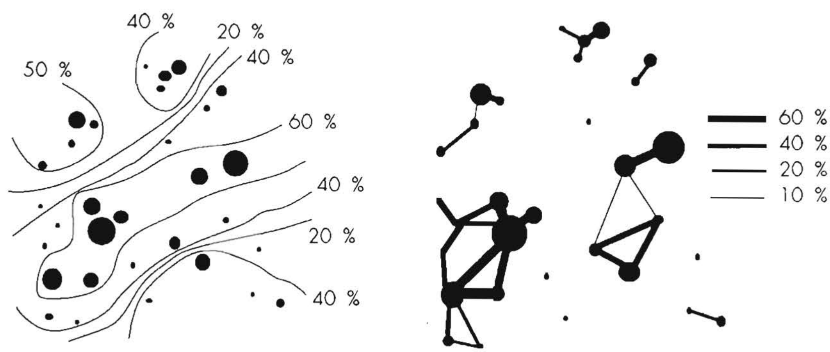


Figure 7.4 Two methods of displaying clusters on a circle map: the isoline method (left) and the Delaunay triangle method (right). In the isoline method, the curved lines surround areas in which circle symbols are grouped in clusters by the same percentage of subjects. Each of the curves represents the responses of a specific percentage of the subjects. In the Delaunay method the line width represents the percentage of subjects that regard the connected stimuli as belonging to the same group.

The frequency of occurrence at which individual circles are perceived as being part of one or the other cluster are not counted, but how frequently *adjacent elements* are regarded as belonging to one and the same cluster. Pursuing this approach to the symbols in a rectangular matrix would mean having to consider all pairs of adjacent cells in the horizontal, vertical and perhaps even diagonal directions. The analysis of the perceived clusters is not further elaborated in the Slocum and Gilmartin study. Apart from counting how frequently circle pairs are clustered and marking the number of times on the map by means of graduated line widths between Thiessen neighbours, no methods for a further analysis are proposed.

A 1983 study by Slocum can be regarded in this respect as a sequel to the study by Slocum and Gilmartin. The design of this study was comparable to that previously discussed, subjects had to draw lines around visual clusters on a number of graduated-circle maps (a total of 14 different maps showed the distribution of various socioeconomic and agricultural phenomena). In addition, the perceived clusters had to be rated on a 6-point scale ranging from extremely poorly defined clusters (rating 1) to extremely well defined clusters (rating 6). For each of the individual subjects (n=61), pairs of circles were only considered clustered if the pair was part of a region on the map around which a line was drawn in the first part of the task and this cluster was given a rating of four or more in the second part. As in the Slocum and Gilmartin study, the degree of clustering was defined as the percentage of subjects who gave the above response. Pairs of stimuli were determined by Thiessen polygons and Delaunay triangles.

Three hypotheses, based on both intuition and previous work of psychologists and cartographers were developed to explain the location of perceived clusters.

First hypothesis

The first hypothesis was that readers cluster circles that are in close proximity to one another. This hypothesis is based on the Gestalt principle of proximity. A ratio of the edge-to-edge distance (b_{ij}) and centre-to-centre distance (c_{ij}) of the individual circles in all pairs (i,j) was calculated as a simple measure of proximity ($p_{ij} = b_{ij}/c_{ij}$). This measure was then standardized for the map environment, resulting in

$$p'_{ij} = \frac{p_{ij} - \bar{p}_j}{s_j} \qquad [7.1]$$

where p'_{ij} is the standardized proximity measure for the *i*th pair on the *j*th map, p_{ij} is the simple proximity measure computed above for the *i*th pair on the *j*th map, $\bar{p}_j$ is the mean of all ratios on the *j*th map and s_j is the standard deviation of the ratios on the *j*th map.

Second hypothesis

The second hypothesis was that readers cluster circles that are similar in size. This hypothesis is based on the Gestalt principle of similarity of size. The ratio of smallest to largest symbol in each pair of Delaunay neighbours was calculated, resulting in a set of ratios ranging from 0 (when one of the counties on the map contained no circle symbol) to 1 (the symbols of two neighbouring counties are of equal size). This similarity measure was also standardized comparable to the above-mentioned proximity measure, giving

$$e'_{ij}=\frac{e_{ij}-\bar{e}_j}{s_j} \qquad [7.2]$$

where e'_{ij} is the standardized equality or similarity measure.

Third hypothesis

A third hypothesis concerned the immediate environment of clustered circles. Map readers were thought to cluster circles within a region having a high ratio of circle area to white background. A pair having a relatively large sum of diameters (compared to the diameters of the circles in the environment of this pair) would have a higher percentage of subjects clustering it. This measure can be compared to the Gestalt principle of figure and ground. In the analysis of this measure both the local map environment and total map environment have to be considered. Inclusion of the total environment was needed to account for the effects of the scale of the map. A map X that contains circles that, on the average, are larger than those of another map Y does not necessarily contain a larger number of clustered pairs. If, however the sum of diameters is not considered relative to the total map environment, then maps on larger scales would necessarily have larger sums of diameters and the likelihood of clustering would be correspondingly greater. The local environment was considered on the basis of a two-step procedure. First the diameters of all circles within a visual circle were summed. This visual circle measured 1.5 inch in diameter and was centred on the midpoint between each pair of Delaunay neighbours. The visual circle represents an area that can be seen with a high degree of clarity during fixation at a normal reading distance. In the second step, the sum of diameters within the visual circle was divided by the proportion of the visual circle within the map boundary. This second step in the environment justification was deemed necessary in order to counteract the possibility of a relatively small number of perceived clusters near the boundary of the map. The sum of diameters for visual circles that partly cross the map boundary will be proportionally less than those that are completely within these boundaries. The effects of local and total environment are combined by dividing the local environment measure by the sum of diameters for all circles on the map (Slocum, p. 66). The resulting figure-ground measure for a circle pair (f_{jk}) becomes

$$f_{jk} = \frac{\dfrac{\sum_{i=1}^{n} d_i}{p}}{\sum_{i=1}^{N} d_i} \qquad [7.3]$$

where d_i represents the diameter of circle i, n is the number of circles within the visual circle, p is the proportion of a visual circle within the map boundary and N is the number of circles on the map. The visual circle is centred on the midpoint between stimulus j and stimulus k.

A logit analysis was used to test the three hypotheses. Results showed that all three measures were significant at the 0.001 level, indicating that each measure was explicitly related to the perceived clusters. The squared correlation coefficient[1] for the figure-ground measure was found to be as large as 0.26, for the measure of proximity a ρ^2 of 0.15 was calculated and the logit analysis of the similarity measure revealed a ρ^2 of only 0.01. That even this low ρ^2 value was found to be significant is largely due to a very large sample size of 2184 stimulus pairs. In this respect we note that Slocum probably added all 156 stimulus pairs or Delaunay neighbours from all 14 maps to calculate only one single ρ^2 for each of the three measures.

Because the model used by Slocum gives a reasonably good description of the visual properties involved in the perception of spatial patterns and it resulted in a 92% correct prediction of all circles as clustered or unclustered, we will use a similar design in the following experiment, in which we try to

[1] A ρ correlation was calculated for all three hypotheses, with $\rho 2 = 1\text{-}[L^*(B)/L^*C]$. $L^*(B)$ is the maximized log likelihood for each of the single explanatory variables and $L^*(C)$ is the maximized log likelihood for the model with the best constant. The $\rho 2$ represents the degree to which the hypothesized model improves upon the model with only a constant.

determine the influence of the above-described measures on the perception of clusters in a graphical matrix.

7.3.1 Experiment 1

Stimulus material

Seven different matrices were used in our experiment. The size of each of the matrices was 20 rows by 20 columns. The matrix cells were filled with circle symbols. For bar symbols, the edge-to-edge and centre-to-centre distances are different for horizontal and vertical stimulus pairs. As this almost certainly affects clustering, we will, for the time being, use only circle symbols. The sizes of the individual circles were equal to those used in the size estimation experiment discussed in the preceding chapter, and were randomly determined for each of the 400 cells of the matrix. Because of this randomization, we may expect an amount of blackness that is about equal for the 7 matrices. The context of this experiment differs somewhat from that of Slocum. In a rectangular matrix it is much easier to define the relevant pairs of stimuli. All pairs of cells which are neighbours in horizontal and vertical directions can be regarded as Delaunay neighbours. In addition to these horizontal and vertical pairs, it is also possible to cluster diagonal pairs, so that these also have to be included. In this way, a total of 1482 different pairs can be distinguished on each of the maps.

Procedure

Seven cards, each containing a different matrix, were compiled into a booklet. The order of the cards in the booklets was counterbalanced across subjects. Subjects were asked to draw lines around "visual clusters", groups of circles that appear to belong together and form a visual unit. This instruction is similar to those in previously discussed studies (Jenks, 1975, Slocum and Gilmartin, 1979, Slocum, 1983). When the subjects were finished with this first part of the task, they had to return to the first matrix and rate the previously drawn clusters on a five-point scale (ranging from 1: poorly defined cluster to 5: well defined cluster). Subjects were not allowed to indicate new clusters on the maps while performing the second part of the task. By separation of the two tasks we expected to get a more consistent rating of the specific clusters across the 7 cards.

In order to be considered as clustered, a circle pair had to meet two requirements. For an individual subject, both stimuli in the pair were part of a group around which a line was drawn in the first part of the task and this group was rated with a value of three or more in the second part. For the entire set of subjects, the degree of clustering for a specific circle pair was the percentage of subjects that made the above response.

Subjects

Twenty-two subjects participated in the experiment. All subjects were students of the Faculty of Architecture, Building and Planning of the Eindhoven University of Technology. Subjects were paid for participation.

Results and Discussion

Analogous with the study by Slocum (1983) percentages clustered were transformed to a logistic curve. In an ordinary least-squares regression analysis the effects of proximity, similarity and figure-ground on the perception of clusters were analysed.

Results showed that although the figure-ground measure gave the best prediction of perceived clusters, there was a considerable difference in the r^2 values between the different matrices (see Table 7.1). The r^2 of this figure-ground measure, for instance, ranged from 0.13 to 0.35. All r^2 values of Table 7.1 are highly significant ($p<0.001$) which is partly a function of the large sample size[2] of 1482. The overall r^2 including all stimuluspairs of all 7 matrices was 0.24. The same overall measures for similarity and proximity were respectively 0.20 and 0.19.

Table 7.1 Regression values (r^2) for each of the independent variables and for a combination of all three measures. All r^2 were significant at 0.001.

		Matrix number						
Independent variables		1	2	3	4	5	6	7
figure-ground	(A)	0.31	0.18	0.27	0.35	0.13	0.25	0.23
proximity	(B)	0.22	0.16	0.19	0.24	0.15	0.18	0.21
similarity	(C)	0.18	0.16	0.19	0.25	0.20	0.16	0.29
total	(A+B+C)	0.40	0.30	0.36	0.46	0.31	0.35	0.40

2 Magnitudes of $\rho 2$ values are much lower than r2 values obtained for the same data. Slocum (p. 64) mentions that the highest $\rho 2$ value of 0.28 that he obtained for the data in his experiment corresponded to an r2 value of 0.68 for the same data. When we consider that differences between these two types of correlation coefficients can be of this magnitude, the r2 values in Table 7.1 are, although significant, rather low.

Although Slocum does not separately analyse the various maps that he used in his experiment, he makes it appear that the correlation values did not differ between the various maps. In our experiment some clear-cut differences in correlation coefficients between maps were found.

In order to explain the differences in the explanatory power of the independent variables between maps, we visually studied the different maps presented to the subjects. This inspection showed that, in spite of randomization of circle sizes, the circles in some of the maps seemed to be somewhat better organized. Even with a randomization of circles of various sizes, subjects appear to recognize clusters and, the same ones, at that. For example, this can be disclosed by measuring the correlation (r^2) between the number of subjectively perceived clusters in a matrix and the corresponding "total" correlation value, A+B+C of the same matrix (as shown in Table 7.1). This correlation was 0.84, which means that the explanatory power of the independent variables is related to, and increases with, the number of perceived clusters. Clusters apparently become easier to predict when the circle symbols are better organized in a matrix.

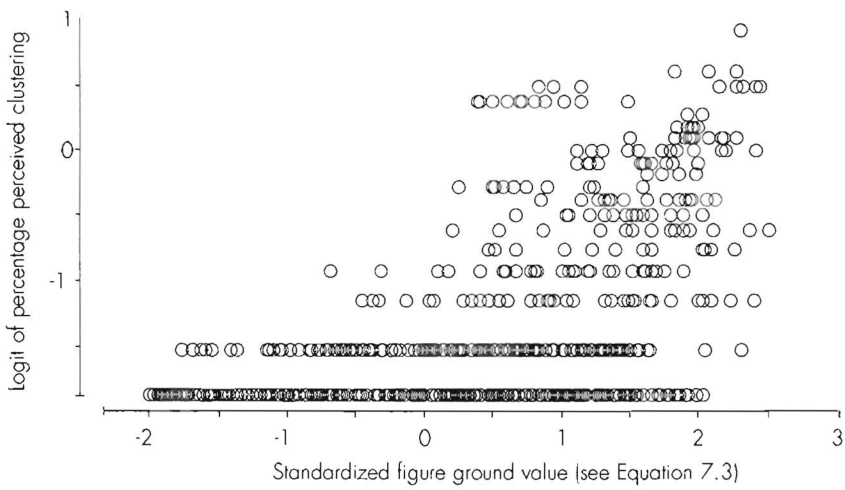


Figure 7.5 Plot showing the distribution of logit values of percentage perceived clustering to the values of one of the independent variables: the figure-ground measure (equation 6.3). Displayed data are from one of the matrices (matrix number 4 of Table 7.1) presented in the experiment. Circle symbols in this figure represent the 1482 stimulus pairs of the matrix.

A second observation is on the method of Slocum. He adds the different cards and estimates only one average ρ^2 value for each of the independent variables. Of the total number of incorrectly predicted circles as clustered or unclustered in the Slocum study, 37% were located on two of the 14 maps, whereas another 8 maps contained only 24% of these wrongly predicted circles. This unequal distribution of incorrectly predicted elements is in line with the differences in r^2 between matrices found in our experiment and makes Slocum's method of simply adding all stimulus pairs of different maps highly questionable. Thus, the ρ^2 might vary more between maps than Slocum expects.

When we look at the scatter plot in Figure 7.5 that displays the relation between logit values on the y-axis and the figure-ground values on the x-axis we note two things (in this figure, results of only one of the matrices, namely matrix number 4 are shown). First, we can see that all stimulus pairs, represented by circles in the figure, that are perceived to be clustered (by a large number of subjects) are concentrated in the right half of the figure. This means that the figure-ground measure (equation 7.3) is definitely related to the perception of clusters. If we were to use the figure-ground measure to predict the location of the perceived clusters, by requiring a minimum figure-ground value, we would not miss too many of those actually perceived. The second conspicuous aspect of this figure is that even in the right half of the display there are many stimulus pairs that have a low logit value. A prediction of the perceived clusters would result in a large number of stimulus pairs that, in fact were not visually recognised as clusters. The same phenomenon applies to the measures of proximity and similarity and to a less extent also to the total value (A+B+C in Table 7.1) that includes all three independent variables.

To explore these results further, some computer print-outs of the original matrices were made in which lines were drawn between all stimuli of the stimulus pairs that were perceived to be clustered by 25% or more of the subjects[3]. In a similar set of print-outs, all stimuli of the pairs that were predicted to be clustered, as based on the figure-ground measure, were connected. When the stimulus pairs of these two drawings are visually compared, two remarkable differences can be noted.

1. A large number of the groups of circle symbols that are predicted to be clustered, but perceived to be unclustered, contains only a small number of elements. For matrix nr. 4, a total of seven of the predicted clusters

3 The value of 25% is based on a visual inspection of the plots in which logit values were plotted either to one of the independent variables or to an aggregated value of all three independent variables. An example of such a plot is shown in Figure 7.5.

consisted of only 2 or 3 elements. The average number of elements in the perceived clusters is 7.5.

2. The perceived clusters generally have a regular shape, whereas a large number of the predicted clusters shows all kinds of irregular protrusions. This is illustrated in Figure 7.6. In this figure a small part of matrix number 4 is shown. In the left part of this figure the stimulus pairs that were perceived to be grouped assume an almost rectangular shape. The predicted pairs are shown in the right-hand part.

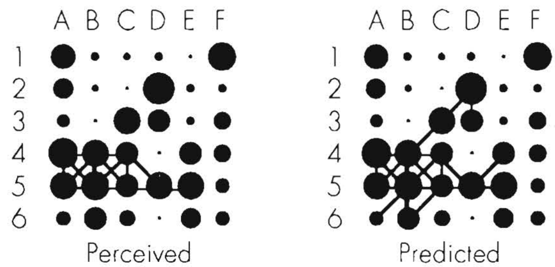


Figure 7.6. A comparison between perceived stimulus pairs and pairs predicted by the figure-ground variable (equation 7.3). Lines that are drawn between the circle symbols indicate that these symbols were either predicted to be clustered or perceived to be so by 25% or more of the subjects.

Stimulus pair A6-B5 is one of the predicted pairs, but A6-A5 and A6-B6 are not considered to belong together. The same effect can be seen with respect to stimuli E4 and E5. Although each of them is connected with D4, they are not connected with each other. In the string B4-C3-D2-D3, each of the elements is only connected with the following one, but this string is not continuous. In the perceived clusters these irregular-shaped clusters seldom occur. Some of the continuous or less continuous strings of elements even linked a number of more coherent clusters. Thus large, irregular-shaped groups of symbols could be seen.

These two apparent characteristics of wrongly predicted pairs make them traceable. Introduction of the correct control variables would diminish the occurrence of such groups and would reinforce the predictive power of the model. The effect of irregularly shaped clusters could be diminished by tracking the contour line of the cluster (and examining its convexities) or by counting the number of connections between a specific element of the group and its other elements. Criteria could be drawn up governing the minimum number of required connections before an element could be considered as part of a cluster. These criteria would have to take account of the position of the stimulus within the cluster; as an element is surrounded by a larger number of other elements, it also needs more connections in order to be regarded as a fully qualified member of the cluster. The number of circles in a cluster seems to be a second useful control variable. To introduce both these variables would mean that requirements of proximity, similarity and figure-ground are not restricted to the immediate Delaunay neighbour(s) of a specific circle stimulus but would also include other elements which, though not Delaunay neighbours, are nonetheless located in the immediate environment.

Some observations are called for on the comparison of our experiment with that of Slocum. Slocum states that 82 percent of all neighbouring pairs were chosen in either the 0-25 or the 76-100 percent category. The results of our experiment are even more extreme, only 530 stimulus pairs (5.1%) were clustered by more than 25% of the subjects and given ratings of 3 or more, whereas all other pairs fell within the 0-25 percent category.

Another major difference between the cartographical maps of the previously discussed studies and the matrix in our experiment is the kind of data that are shown. Actual socioeconomic and agricultural phenomena were shown on cartographical maps. In the matrix, on the other hand, the sizes of circles were determined randomly. It is not illogical to assume that at least some of the phenomena shown on the map are inherently less uniformly distributed across the counties of a state, which would be reflected in more and larger coherent groups of similar-sized stimuli on the map and lead to easier perception of these groups. In the matrix, the occurrence of large, coherent groups is less obvious.

On examining the particular maps of subjects, we noted that some subjects had drawn lines around similar-sized but very small symbols, around strings of adjacent circles that showed a continuous increase in size from one end of the cluster to the other, or around groups of circles in which a certain type of symmetry was to be seen. These stimuli might seem to belong together from an aesthetic point of view, but were not really intended to be clustered. Obviously, our intention was not followed properly by these subjects. Another possibility is that they naturally tended to employ other criteria of grouping.

The nonoptimal explanatory power of the independent variables definitely calls for an increase in the number of independent variables. To be more specific, how do the stimulus pairs perceived to be clustered in the right-hand part of Figure 7.5 differ from the stimulus pairs not perceived to be clustered in the same part of the figure? The previously mentioned "minimum number of circles in a cluster" and "contiguity of the shape of a cluster" could prove to be rewarding additions.

7.4 Rating of structure

The second object of this chapter involves the processing of information at the overall level. We can divide this problem into two questions, one regarding the perception and subjective interpretation of the overall distribution of elements in the matrix (or in brief, the order of the matrix) and the other concerning an objective measure about the organization of elements in a matrix.

Objective measure

To start with the second issue, we could ask ourselves the question whether there are any standard statistical or other objective mathematical measures that give an indication of the subjective ordering of different elements in a rectangular whole. The structural information theory (Leeuwenberg, 1971, van der Helm, 1988), for example, aims at describing complex visual patterns in a concise coded format containing as few elements as possible. The (objective) minimum number of elements needed to describe a graphical matrix and the subjectively rated orderliness could very well be related. In order to attain this code, the structural information theory introduces a number of coding rules (based on previous work of, among others, Attneave, 1957 and Garner, 1974). Experimental research has shown that an important role is played by three types of regularity, that is iteration, symmetry and alternation (van der Helm, 1988). These types of regularity, however, do not fit in with our interpretation of the concept of structure.

We have to bear in mind that our graphical matrix represents a number of characteristics that are plotted to a number of objects. When one or both sets (objects and/or characteristics) are ordered, this means that characteristics or objects having the same values are juxtaposed within the matrix. The amount of structure or order is, on that account, directly related to the number of elements that are classified in groups, where there is as little variance as possible between the elements in a group. The regularity types of symmetry and alternation of the structural information theory are therefore not in accordance with this interpretation of the structure concept. Moreover, the structural information theory does not take the size of differences into account. Elements are either the same or different. In our interpretation, the size of differences matters positively. The intensity of the relation between two graphical elements shows a continuous decrease as the difference increases.

A more valid measure of ordering is given by the statistical distance measures. The Euclidean and the city-block metrics, for example, give an indication of the distance between two projected points on a two-dimensional plane. A comparable measure can be introduced in a rectangular matrix. If the distance between two adjoining elements (i,j) and (i,j+1) in the horizontal direction or (i+1,j) in the vertical direction is defined as ($n_{i,j}$ - $n_{i,j+1}$) or ($n_{i,j}$ - $n_{i+1,j}$), where ($n_{i,j}$ is the frequency or size of the element in cell i,j) a measure of distance (d) of a specific configuration of elements in a matrix is obtained by

$$d = \sum_{j=1}^{k-1} \sum_{i=1}^{r} |n_{i,j} - n_{i,j+1}| + \sum_{i=1}^{r-1} \sum_{j=1}^{k} |n_{i,j} - n_{i+1,j}| \qquad [7.4]$$

An index of order (o) can then be defined as

$$o = \frac{1}{d} \qquad [7.5]$$

In equation 7.4, the first component of the right-hand part expresses the total distance between columns, the second component indicating the total distance between rows. The size of the specific elements is equal to n_{ij} which normally stands for the frequency in the *ij*th cell. The variable r equals the number of rows and k is number of columns.

This measure includes only the differences between neighbouring elements in the horizontal and vertical directions, which is in accordance with our interpretation of the concept of order[4]. Whereas the distance functions originally give an indication of metric distances between two projected points on a two-dimensional plane, in our context this measure expresses

[4]In the pattern-recognition experiment, neighbouring elements in a diagonal direction were also reckoned with in setting up the standardized similarity and proximity measures. By analogy, the order equation can be extended by inclusion of these diagonal pairs. As the centre-to-centre distance between diagonal pairs is greater than that between horizontal or vertical pairs, these pairs should be weighted accordingly. These weights could, for example, be calculated using the Euclidean distance measure or the city-block metric (Everitt, 1978).

the areal size differences between two juxtaposed circle symbols in a matrix. A disadvantage of this measure is that it is tied to the matrix at issue. Different matrices will return different values of ordering and the measure does not permit direct comparison of matrices that contain different sets of elements. This disadvantage continues when we are in search of the best solution or configuration. Although the orderliness indicator accepts the better ordered matrix pointed out in a series, it cannot tell whether this relatively best matrix is also absolutely the best one attainable.

Subjective measure

A first impression of the subjective interpretation of order is obtained through an experiment in which a set of cards had to be ordered. A number of subjects had to align a set of cards on the semantic differential "ordered - unordered" (rank them according to the amount of clustering of the symbols in the matrix). Comparing and judging matrices is a task analogous to an evaluation of similarities between cartographical maps presented in pairs or triads. A large number of studies has shown that subjects are very capable of performing this type of task (Lloyd and Steinke, 1976,1977, Muller, 1975, Olson, 1975a). In a second experiment, subjects had to rate the difference in orderliness of two pairwise presented matrices on a continuous scale. Results of these experiments can be used, not only to test the generality of the concept of structure, but also validate the measure of "order" used in the analysis.

7.4.1 Experiment 2

Stimulus material

Three series of cards were presented to six subjects in this experiment. Each card contained a drawn 25 · 25 matrix with a number of cells filled with dot-like symbols. An example of a card, as presented in this experiment, is shown in Figure 7.7. Two series (A and B) consisted of 8 cards, the third series (C) had 10 cards. Circles of two different sizes were used in series A, that is, 125 small dots and 102 large dots. In series B the circles were of 3 different sizes, with 44 small, 60 medium and 79 large dots. Four different categories were used in series C, with each category containing 125 circles. Cells that didn't contain a circle symbol were left empty. All cards within a series contained the same circle elements, the only difference being the location of the circles in the matrix. The different configurations (= cards) within a series were obtained through a number of consecutive exchanges of a variable number of rows and columns of the same basic matrix. Statistically, this means that all cards within a series have the same χ^2 value.

Procedure

Subjects had to order the cards or matrices of each series according to the "orderliness" of the circles in the matrix. In this instruction, orderliness was defined as the degree to which elements of similar size appeared to be grouped.

Subjects

Six male students of the Faculty of Architecture, Building and Planning participated in this experiment. The experiment was combined with another in which circle symbols had to be interactively grouped. This grouping experiment will be discussed in the next chapter. The card-sorting task was performed twice; preceding and following upon the grouping experiment. Each of the two sessions took only about 5 minutes; the time interval between the sessions was two to three hours.

Results

The association or correlation between the ranking of the series by the different subjects was measured by Kendall's coefficient of concordance, W (Siegel and Castellan, 1988, p.262). Results of the 12 sets of ranking (the two sessions of each individual were regarded as results from different subjects) showed coefficients of 0.98, 0.97 and 0.95 for the three different series of respectively 8, 8 and 10 cards. All three values were highly significant ($p<.001$). Separate measurements of the sessions before and after the grouping experiment gave similar results. Comparing (the separate rankings of) the two sessions for all three series and all subjects showed significant results for all comparisons, with r_S (Spearman rank-order correlation) values ranging from 0.83 ($p<.05$) to 1.0 (Table 7.2). Finally, correlation coefficients between every subjective ranking and an objective ranking (based on equation 7.5) were calculated (also r_S). All coefficients were significant, with values ranging from 0.78 ($p<.05$) to 1.0.

Table 7.2 Rank-order correlation coefficients (r_S) between the two sessions (range of the individual coefficients in the second and their mean in the third column) and between the objectively defined measure and the subjective performance (fourth and fifth columns)

	between-session reliability		reliability between obj. and subj. measure	
series	range	mean	range	mean
series A	0.83 - 1.00*	0.95	0.78 - 1.00*	0.97
series B	0.95 - 1.00**	0.98	0.95 - 1.00**	0.98
series C	0.92 - 1.00***	0.98	0.95 - 1.00***	0.98

* all $p<0.05$; ** all $p<0.005$; *** all $p<0.001$

Discussion

The results of this experiment indicate that the concept of orderliness of graphical circle elements in a matrix is

- a generally accepted concept which can be confirmed by the high degree of "between subjects" agreement.
- a consistent phenomenon, which is confirmed by the high degree of "within subjects" agreement (between sessions).

The objective measure based on the "distances" between the elements in the matrix (equations 7.4 and 7.5) seems to give a fairly accurate description of the underlying concept and results in a high correlation between the objective statistical measure and the subjective performance.

This last point merits an observation. In the validation of the objective measure we related an ordinally scaled variable (order number of a card in a series) with a variable that was originally at an interval or ratio level (objective value of orderliness). After comparing two cards in a series, we can only conclude that one of them was, subjectively considered, more ordered, whereas the objective measure allows a more exact specification of the difference or ratio in orderliness between the two cards. If the subjective opinion were to be measured at a higher level of measurement, the validation of the objective criterion might become stronger. This was done in the following experiment.

7.4.2 Experiment 3

Stimulus material

In this experiment, two graphical matrices were presented at the same time on a computer screen. The same 3 series of cards were used as in the preceding test. One of the matrices (the reference matrix) was the same throughout the whole series. The relative orderliness of the reference was moderate, so that some targets were better ordered, while others had lower values of objective arrangement. To prevent too easy recognition of the reference matrix in the pairwise presentations, the reference matrix was randomly alternated with its mirrored or rotated image, on the hypothesis that these operations did not change the perception. In addition, the position of target and reference matrix in the display (right or left) was randomized. A target matrix was displayed beside the reference matrix. Target and reference were always of the same series.

Procedure

The task of the subjects was to mark the difference in orderliness between the target and reference matrix. For that purpose they had to mark two points on vertical lines, one point and vertical line for each of the matrices. A higher position on the scale indicated a better ordered matrix. Subjects were instructed to choose the marking points in such a way that these expressed both orderliness of the individual matrices as well as the difference between them. The lines were projected midway between the reference and target (see Figure 7.7).

The rating of differences in amount of structure between matrices in a series is a more demanding cognitive task than rank-ordering the same series of matrices on their amount of structure. As compared to the rank-ordering task, rating of differences requires an additional determination of the exact magnitude of the difference and a subsequent translation of this magnitude into a numerical value, or here, to a difference in height on yardsticks. In each trial of the rating task, however, only one pairwise comparison has to be made and decisions for each pair of matrices are independent, whereas the rank-ordering task requires a multitude of related comparisons (Lloyd and Steinke, 1976, 1977). In addition to the previously discussed distance values, the response time of the trials was also recorded. Response time is defined as the time between the presentation of the matrices on the screen and the second response of the subject on one of the yardsticks. A new trial started and another pair of matrices was presented on the screen three seconds after this second response had been given. There were no time restrictions in this self-paced task.

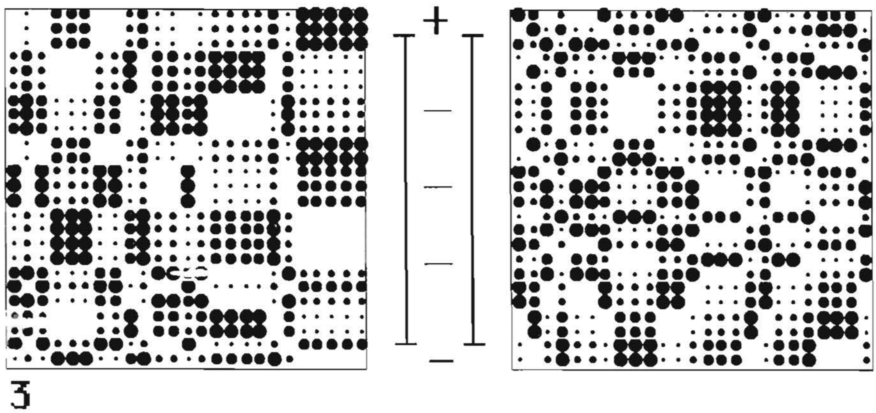


Figure 7.7. Example of a pair of matrices with yardsticks and trial number as presented on the screen during the experiment. Matrices had to be rated by pointing to the yardsticks at lengths proportional to the orderliness of the matrices.

Subjects

Six subjects participated in this experiment, 5 male and 1 female. All were students or staff members of the Faculty of Architecture, Building and Planning. The three series were presented sequentially and were counterbalanced between subjects. Two of the subjects performed the task twice, at an interval of at least one day. The other subjects participated once. Performance of the task took 10 to 15 minutes. Subjects were not paid for participation.

Results

In the analysis, the relation between two subjective and two corresponding objective values were calculated. One set of values consisted of the objective and subjective *differences* in orderliness. Objective difference was estimated by subtracting the orderliness of the reference and the target matrices, calculated according to equation 7.5. The subjective difference was the simple difference in y-value of the two points indicated on the vertical measuring staffs. The other set contained the objective and subjective *ratios* in orderliness which were estimated in the same way. Correlation coefficients (Pearson r) for the differences and ratios are given in Table 7.3.

Table 7.3 Correlation coefficients between objective measures and the subjective performance for their differences and ratios. Range of individual coefficients with their mean, median and standard deviations are shown.

	differences				ratios			
series	range	mean	median	St dev	range	mean	median	St dev
series A	0.91 - 0.96	0.94	0.95	0.02	0.73 - 0.97	0.88	0.90	0.09
series B	0.77 - 0.95	0.89	0.90	0.06	0.74 - 0.93	0.87	0.89	0.07
series C	0.83 - 0.97	0.91	0.91	0.05	0.79 - 0.98	0.87	0.86	0.07

The objective differences and ratios between target and reference matrix were closely related to their corresponding subjective counterparts. The value indicating the difference between matrices seems to be somewhat more consistent than the ratio indications when subjects are compared. This is expressed by a shorter range and a lower variance of the first-mentioned correlation coefficients. Within-subject consistency was calculated by correlating the subjective responses of both sessions. The two subjects that performed the task twice turned out to be fairly consistent in their ratings, as correlations for the different series ranged from 0.92 to 0.95 for the differences and 0.89 to 0.96 for the ratio measure.

Finally, the relation between response time and the actual differences in orderliness was calculated between target and reference matrix. Correlations varied considerably between subjects and series and most of them failed to reach significance, but on the average, a slight negative tendency could be noted. For the actual differences, the correlations with response time ranged from +0.13 to -0.62, yielding an average of -0.28 (St. Dev., 0.19). For judged differences, the range was from +0.58 to -0.21, with an average of +0.23 (St Dev 0.22). Greater differences between the pairwise presented matrices generally resulted in a slightly (n.s.) shorter response time.

7.5 Conclusion

The results of the experiments on recognition of patterns and order are very promising. Although the variables of proximity, similarity and figure-ground are still not sufficiently elaborated for an optimal prediction of perceived clusters in a graphical matrix, they appear to provide an appropriate starting point for a more extensive model. Inclusion in the predictive model of some perceptively distinguishing characteristics of these clusters, such as the number of circles in a perceived cluster and their regularity of shape, would positively increase its explanatory power. Coherent groups that contain only a few elements are generally not perceived as a cluster. The same is true for discontinuous strings of circle symbols or those elements of a group that can be considered as irregular protrusions. In order to include the variables "number of elements in a cluster" and "contiguity of cluster shape" we have, at least partially, to leave the stimulus pair or Delaunay neighbour as starting point in the sampling network. Next-door neighbours and perhaps even next-to-next-door neighbours of each stimulus in the matrix have to be taken into consideration. To give a more detailed account of the influence of these variables on the subjective perception of clusters, further research on the variables themselves and their relation to those previously discussed is certainly needed.

As regards orderliness in the distribution of circles of various size in a matrix, there is strong consistency in its evaluation or rating, both in time and as between different subjects. A simple objective difference between two points, based on geometric distance functions, shows a strong correlation with the subjectively noted order. These two consistencies make this method broadly suitable for demonstrating and rating improvements in the structure of tabulated data.

In practice, the results of the experiments described in this chapter open up avenues for the graphical presentation of data tables and warrant the use of the graphical matrix for this purpose. Intended groups of circles will be generally perceived as such by the map reader and tables in which data are more coherently organized will also be perceived to be more structured, and be rated accordingly. It should be noted that even in matrices in which graphical symbols were assigned at random to the matrix cells, as in experiment 1, subjects appear to mark clusters. A large part of these clusters are, moreover, identified by a preponderance of subjects.

Once more, the main advantage of graphical presentation of data tables is the immediacy with which their contents are revealed on the intermediate and higher level of information processing. Graphical presentation of data tables by means of the graphical matrix could save much time for people scrutinizing large numerical tables in search of patterns and order.

Chapter 8 Construction of patterns

8.1 Abstract

The interactive aspect of the reorderable matrix is further elaborated in this chapter. In an experiment, subjects were required to structure a collection of circle elements that had been mixed together before presentation on a visual display terminal. Two aspects had to be considered in this structuring task. The first was to place the scattered circle symbols in groups or clusters, the second involved joining the developing groups along a continuously increasing or decreasing line (function) within the matrix.
In the easier experimental conditions matrices were presented which could be reconstructed into a number of clear-cut rectangular groups with little overlap between them. They were all easily dealt with by the subjects, even though the number of operations needed to achieve the optimum differed widely. With the more difficult matrices, results were more divergent.
As not all of these matrices were solved, the number of operations and the rated end results differed between subjects and specific experimental conditions. Study of the structuring process itself showed that the best results were obtained when both ordering and association aspects were under review throughout the process. The number of operations needed to reach practically optimum end results depended on the complexity of the actions carried out by the diverse subjects. In this strict experimental context, some standard clustering algorithms proved useful in putting the subjects on the right road.

8.2 Introduction

The results of the experiments that were discussed in Chapter 7 yielded two important results.
First, it was found possible to roughly predict groups of elements in a matrix that map readers actually perceive to be clustered. The second major result refers to the general level of information processing. At this level, the perceived orderliness or amount of structure in a matrix proved to be very consistent. At the same time, the concept could be accurately described by a simple model. The implication of these two findings is that the graphical matrix offers some very interesting possibilities as a tool for the *presentation* of tabulated data.

On one hand, the display remains close to the original data set in presenting all its individual items. On the other, the matrix proves to be suitable for fast and easy processing, not only of individual elements but also of larger chunks of information, coherent clusters of symbol elements or even the complete picture (read matrix). The matrix can therefore be regarded as both a large collection of small graphical symbols and as one visual image with possibly some conspicuous properties. The complete image and the clusters, as its outstanding parts, are easily recognized and rated as such on the semantically differential concept of structured - unstructured.

In order to be useful in the *analysis* of information, the graphical matrix has to meet a second major requirement, that is the possibility of actively organizing data. Multivariate data, gathered in order to substantiate decision making in architectural and planning problems, mostly originate in a large unorganized collection of information items. The data are not always specially collected for a specific investigation, but can be (partly) copied form existing data bases. We can think of possible information sources, such as the geographic location of building sites, demographic information on town development and knowledge about city traffic. In these sources objects and their characteristics, for example, may be originally ordered alphabetically or chronologically. In order to make decisions that are based on, and verified by, the original information, the original layout of objects and characteristics often have to be reorganized. Before similarities, contrasts and relationships between objects can be brought to light, the information needs to be given a more structured appearance and groups with elements of similar size have to be made conspicuous. Translated into the matrix format, this would mean that we have to rearrange rows and columns of the matrix. If map (matrix) readers could indeed discover similarities, contrasts and relationships within the matrix by active organisation and clustering of its elements, it would be clear that the graphical matrix is not only suitable as a *presentation* tool but also as an *analytical* tool. When the hypothesis "the graphical matrix is a usable tool in the active organisation (relationships) and clustering (similarities and contrasts) of multivariate data" is at stake, two issues arise. The first concerns an operationalisation of the specific task laid on the subjects (the definition of the verbs to organize and to cluster, stated in the preceding sentence), its validation and the evaluation of its results. The second issue concerns the tool or program itself. To test the hypothesis, we have to create a tool that is optimally adapted to the interactive task that it has to execute. Since we do not know in advance what strategies people will use and what specific actions they want to perform, we can, for the time being, only offer them some basic, mathematically legitimate options in the ordering process.

8.3 Operationalisation of the structuring task

Structuring the information in a rectangular data matrix requires us to distinguish two different aspects in the task. The first has to do with grouping or clustering equivalent elements and will be discussed in the next section on orderliness. The second aspect, concerning the overall organisation, moves somewhat more in the direction of interpretation of data and has to do with trying to discover underlying relationships between the two components of the matrix, the objects on one axis and their features on the other axis. This aspect will be discussed in the section on association.

8.3.1 Orderliness

In Chapter 7 the perceived orderliness of elements in a matrix was evaluated by comparing the subjective ratings of matrices to an objective measure. This measure, based on the "distance" between elements, gives an indication of the "degree of grouping" in elements of similar size in clusters. As this degree of grouping is also equivalent to the degree of adjacency of objects and features with similar profiles in the matrix, there is a direct functional relationship between the distance measure and the grouping of equivalent objects and features. One of the purposes of the reorderable matrix is to discover similarities and contrasts between different objects or between various features. As far as this purpose is concerned, our distance measure seems to provide a correct bench-mark in the analysis of authentic multivariate data.

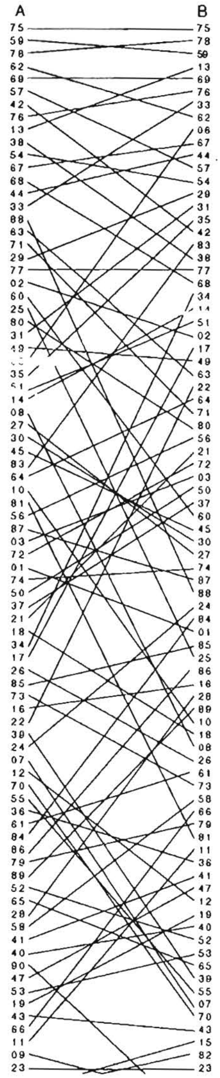
There is, however, one important restriction to this measure, it gives a description or a relative appraisal of a specific configuration of elements in a matrix, but is not suitable for seeking an optimal solution in an eventual search. Although we could theoretically calculate the orderliness of each possible configuration, in practice this method is coarse and time-consuming. We should bear in mind that the number of possible configurations equals the product of the factorials of the number of rows or features and columns or objects ($N=r! \cdot k!$, where r equals the number of rows, k equals the number of columns and N equals the number of possible configurations). With a matrix of 25 rows and 25 columns there are about $2.4 \cdot 10^{50}$

possible and unique ways of ordering the rows and columns. Thus the search space (e.g. Gick and Holyoak, 1979) for large sets of data (long x and y-components) becomes extremely large.
It is also difficult to attain the optimum configuration of the row and column variables by more intricate statistical or mathematical analyses. Analyses normally used with problems related to the present one, often require additional information on the displayed items, such as the number of clusters that are to be expected and the permissible maximum within cluster variance. The shape of the clusters created by different clustering techniques depends on the specific algorithm used. Selection of the right algorithm, in turn, depends on the hypothesis to be tested and the characteristics of the data set itself. If this information were at hand, we could, in a very detailed elaboration and operationalisation of the hypothesis, compare the results of some of the corresponding clustering algorithms to results obtained in a subjective interactive approach. The results of these analyses could be compared with the earlier described measure of ordering. Some of the disadvantages of the statistical approach have been briefly discussed in Chapter 1.

8.3.2 Association

In addition to the ordering of equivalent elements in the matrix, a second aspect of a well-organized table is the way it reveals a tentatively present association between the two components of the table (the x and y-components or the objects and features along the axes). This search for an association that is operative between the whole line of the two components of the matrix and not just between a small number of objects or features, in the form of clusters, is often enacted on the transitional area between analysis and interpretation. In a study on the formalization of the Gestalt rule of proximity in pattern description, van Oeffelen and Vos (1983) make a distinction between pattern recognition at the perceptive and cognitive levels. They note that, whereas perceptive processes encompass the extraction of features and the detection of simple objects, "cognitive-level techniques mostly deal with formal aspects of picture syntax and scene analysis in so far as they are based on symbol structure manipulation". At this level, pattern recognition depends heavily on the availability of knowledge based on past experiences. For, as far as an analysis of data is no longer based exclusively on visual features, but involves external knowledge on the data displayed, that is, when the data are no longer considered as anonymous, the cognitive level of interpretation has been reached. We should take care not to confuse this cognitive level of interpretation with the cognitive (or strategic) *aspects* involved in the actual process of analysis itself. The latter are a description of the interaction of options provided by the (analysis) tool and the formalized intentions and plans of the subject.

An association between the two components of the matrix can be visually represented by linking the discovered clusters along a continuously descending or ascending line. The aim is thus to maximally differentiate objects across the features component and the latter across the former simultaneously. The fictitious example in Figure 4.10 of the present study, on page 62, gives precisely such a clear-cut, but still very simple, linear association between the objects component (columns of the table) and the features component (rows). Three groups were distinguished in both directions, the exterior ones being each others' opposite. A relationship of such simplicity could also be revealed by analysis based solely on visual characteristics of the display or perhaps even revealed by a specific clustering algorithm. The relationships that Theodorescu revealed in his actual data on ionic capitals (Chapter 1) are much more intricate. To discover a combination of a greater number of simultaneous relationships between the complete x-component and large parts of the y-component might require, not only considerable experience with a tool that enables the reordering of data in a matrix, but also of some extrinsic information on the specific objects and features displayed in the matrix.



A graphic "computation" of association

Bertin (1967) describes a graphical way of comparing associated distributions. In one example he numbers and orders all French départements (administrative districts) according to the percentage of the population working 1, in manufacturing industry (A in Figure 8.1) and 2, in tertiary industries (B). He presents graphically the départements as points positioned on two vertical lines, one for each sector of industry. The vertical position of the points on the line corresponds to their order number. Thus, for each of the two sectors, we get a column with the départements ordered according to, respectively manufacturing industrial population and tertiary industrial population. Next, the two columns are placed alongside each other and the specific départements in the two columns of points are connected by lines. The relative number of intersections of the lines (= actual number of intersections divided by the maximum possible number of intersections) gives an indication of the degree of dissociation between the two variables (Figure 8.1). When there are no intersections, the two variables are perfectly positively correlated (which would statistically result in a maximum-rank-order correlation of 1.0); with a maximum number of intersections, the correlation is perfectly negative.

Figure 8.1 Numbered départements, ranked according to population working in manufacturing industry (left column, A) and working in tertiary industries (right column, B). The number of intersecting lines is an indication of the degree of dissociation between the two variables (after Bertin, 1967).

Presenting these data in a matrix yields another manner of reflecting the relationship between two ordered variables. Assume that the column number in the matrix is appointed by the order number on the first variable

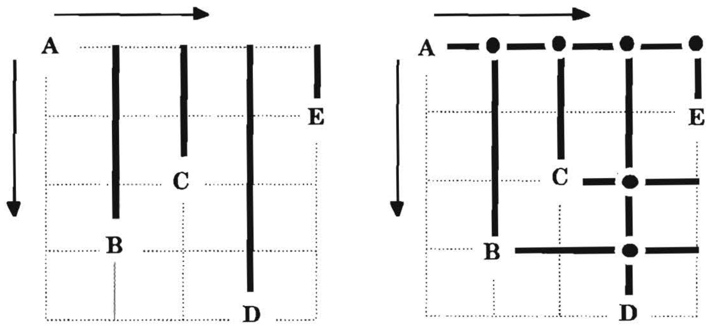


Figure 8.2. Alternative graphic representation of the data in Figure 8.1. The order numbers of a specific département on both variables determine its respective row-and-column number in the matrix (starting at the top left-hand corner and increasing towards the bottom right-hand corner). This layout thus corresponds to that of a table. In this way, each département (denoted by characters) is assigned to one unique cell in the matrix. The number of intersections, indicated by the black dots in the right hand part, indicates the degree of dissociation between the two variables.

(population working in manufacturing industry as the x-component) and the row number as fixed at the order number of the second variable (population working in tertiary industry on the y-component). An example of this method is shown in the left-hand part of Figure 8.2. Carrying out this procedure for every object, each département in our example is assigned to a specific and unique cell of the matrix. Connecting the filled cells with the topmost row and the rightmost column by straight lines again results in a number of intersections. In this type of display, an intersection corresponds to the crossing of a vertical and a horizontal line (see the right-hand part of Figure 8.2). The number of intersections, divided by the maximum number of intersections (=N(N-1)/2) gives a measure of association of the two variables A and B. This measure, as used by Bertin, is in fact a graphically obtained version of Kendall's rank-order-correlation coefficient T.

In this example, only two characteristics of a number of objects (départements) were recorded, namely, percentages of the population working in manufacturing and tertiary industries. As a result, each column and each row of the matrix representation contains only one single element. The graphical matrix offers the possibility of recording, displaying and analysing a much larger number of characteristics of these objects. The same principle and related statistical techniques can also be used in this graphical matrix, which can be regarded as a frequency distribution or cross-classification table with ordered variables. In a cross-classification table, the cells of the matrix contain occurrence frequencies of values X_i of variable X and values Y_j of variable Y. With these cross-classification tables, it is assumed that the values of variable X are ordered in magnitude by their subscripts, that is, $X_1<X_2<...<X_k$. Variable Y is assumed to be similarly ordered, with $Y_1<Y_2<...<Y_r$ (Siegel & Castellan, 1988). In the graphical matrix of the example, this means that a disclosure of the tentative present association consists in ordering both the départements and the characteristics at the same time. Starting at a specific cell in the matrix, the lower ordinal numbers (the ordinal number "5" is lower than the ordinal number "3") need to be located as far as possible in cells that have both a higher column number (more towards the right of the matrix) and a higher row number (more towards the bottom of the matrix). The strength of the association is in proportion to the number of cells (and their contents) that meet these requirements. When the two components along the axes of the matrix are of a nominal or categorical nature, but the position of the specific categories in the matrix is a relevant one, we can treat them as ordinal variables and use the same measures of association as in the ordinal variables.

When we group cells according to contents and chain the developing groups along one of the diagonals of the matrix, we are actually treating the rows and columns as variables of an ordinal scale. As the relative number of intersections also gives an indication of the continuity in the course of one of the components across the other one in the matrix, there is a direct functional relationship between the above-mentioned association principle and the overall association between objects and features. The discovery of association between the two matrix components corresponds to the second purpose of the reorderable matrix, so that our association measure, as it is based on the number of intersections, is held to provide a second valid bench-mark in the analysis of authentic multivariate data.

Just as in the orderliness problem, there is also no fast and easy way to attain the optimum configuration (association) of the row-and-column variables by way of statistical or mathematical analyses. It is only by calculating standard measures of association as, for example the Gamma statistic *G* and Kendall's *Tau-b*, with every possible configuration of the rows and columns, that we would have a means of discovering an optimum solution. Hence, the statistical measures of association are also of a descriptive nature and do not automatically reveal the ideal solution to the problem of association.

As the measures of orderliness and association are both descriptive in nature, we could compare different configurations that are attained in the interactive analysis of a specific matrix problem and order them according to their "rightness".

Methods of analysis

In the following experiment we will use the Gamma statistic G as a measure of association (Goodman & Kruskal, 1954 & 1977; Agresti, 1984; Loether & Tavish, 1974). This is a relatively easily used index that is closely related to Kendall's T. Although this measure does not allow us to find the optimum solution to ordering, it seems a promising one because the method has proved to be appropriate for measuring the relation between two ordinally scaled variables. Moreover, we are interested in a distribution of frequencies providing a continuous increasing or decreasing function and the Gamma statistic G results in a value of 1 as long as there are no "disagreements" in the ordering of both variables.
The gamma statistic G is defined as

$$G = \frac{\#\text{ agreements} - \#\text{ disagreements}}{\#\text{ agreements} + \#\text{ disagreements}} \qquad [8.1]$$

where the number of agreements equals

$$C = \sum_{i<r} \sum_{j<k} n_{ij}\, n_{rk} \qquad [8.2]$$

(C is the number of agreements or concordant pairs, n_{ij} the frequency in the *ij*th cell, r the number of rows, k the number of columns) see Figure 8.3,

and the number of disagreements is

$$D = \sum_{i<r} \sum_{j>k} n_{ij}\, n_{rk} \qquad [8.3]$$

(D is number of disagreements or discordant pairs).

In both equations we will treat n_{ij} as the size of the graphical symbol in the *ij*th cell in the contingency table. Instead of weighting by the frequencies in the *ij*th cells, we will weight them by the sizes of the symbols in the respective cells.

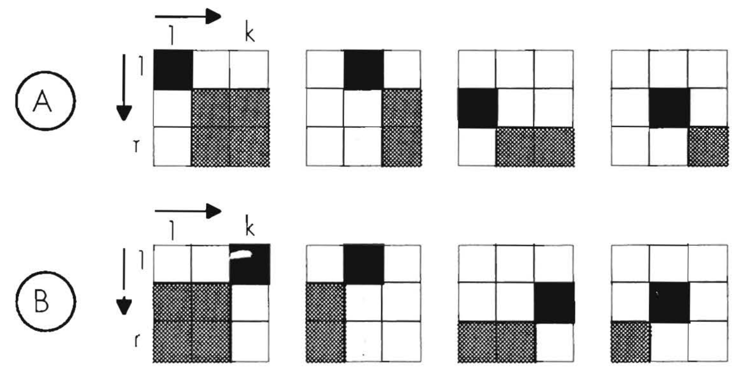


Figure 8.3. Illustration of the method by which the total number of agreements or concordant pairs (part A) and the the total number of disagreements or discordant pairs (part B) of a matrix can be calculated. For each of the four matrices in part A, the number of agreements can be obtained by multiplying the value of cell n_{ij} (black cell) by the sum of the cells for which the row number is larger than i (r>i) and the column number is larger than j (k>j) (gray cells). These four positions are the only ones in which at least one of the cells lies both below and to the right of cell n_{ij}. The same method can be used in part B to calculate the number of disagreements or discordant pairs.

Operationalisation of task aspects

With the distance measure, giving an indication of the degree of grouping on one hand, and the Gamma statistic G as an indicator of the degree of association between the two components on the other, we have two objective, statistical measures that can be used to compute and evaluate all the individual configurations obtainable in an ordering task. These measures can be regarded as an operationalisation of the respective task aspects of clustering and organization, be used to actually test the ordering process, and check whether subjects are able to improve or solve a graphical matrix.
Throughout the process of ordering a matrix in the experimental situation we could test performance and results of the subjects by measuring association and ordering values. After a fixed number of operations, or an agreed period of time, we could calculate and compare the results of these measures. Practically speaking, we can take the line that, so long as subjects are able to improve the distribution of the data by using the interactive ordering method, there is no need to calculate the rejected less-than-optimum distributions.

Qualitative analysis

If subjects appeared to be unable to solve the presented randomized matrices either completely or partially, the ordering process should be examined more thoroughly by study of the actually performed displacements of rows and columns. Such examination could provide some insight into the specific problems that arise while performing this specific task. But even when subjects succeeded in solving these matrices we could learn something by analysing the ordering process. The various techniques or strategies that are used by different subjects and in different practical contexts could be compared and the best ones used for improving the instrument.

8.3.3 Evaluation of process and results.

As the measures of association and orderliness not only give an objective evaluation of the distribution of elements in a matrix, but also cover the two aspects to be found in the purpose of the reorderable matrix, these objective values prove to be valid in the evaluation of the process of ordering and its results. Strictly speaking, it would suffice to record and evaluate the end product of the structuring process attained by a subject performing the task. Depending on the end result, we could conclude whether the specific tool fulfilled our initial intention, which was to structure the information graphically displayed in a matrix, in such a way that it leads to drawing the right conclusions regarding local similarities, contrasts and global relationships.

There nevertheless remain some problems with this proposition. The first major problem concerns the evaluation of the end result. We have already said that, in practical situations using actual data, it is often difficult, if not practically impossible, to mathematically calculate the optimum solution to the structuring problem. In a matrix, this would mean the visually optimum configuration of the graphical symbols. Its practical consequence is that we often do not have an objective "optimum solution" to use as a yardstick with which we can compare practically attained subjective results. This also means that the hypothesis concerning the use of the graphical matrix as an analytical tool needs some readjustment. We cannot test whether the rearrangement of data by means of the reorderable matrix does lead to an optimum solution (except in some very restricted experimental environments), because the optimum solution of practical problems remains unknown. What we can test is whether a rearrangement by displacements of rows and columns leads to a better or a worse configuration. Only in a strict experimental context, where a fictitious ideal data set is randomized before it is presented to the subjects (see Figure 4.10), could we examine the possibility of reaching the optimum solution by using the reorderable matrix. And, even then it must be postulated that the original configuration is also the optimum solution. This also means that improvements and deteriorations in the structure are always relative and can only be measured at an ordinal level of measurement. In addition, performance by different subjects can only be compared within a specific data set and not between initially different sets of data. This leads to the second major problem.

For a subject performing the task it is also difficult, if not impossible, to decide whether or not an attained configuration is capable of improvement. At a specific moment in the process he/she will (has to) decide that the organisation of elements in the matrix cannot be further improved. When, within a strict experimental context, the optimum solution is reached each time the subject decides to stop the process, we could conclude without much ado that the tool suited the purpose, at least in this specific setting. The program obviously offered the possibility of easily finding the best solution to our ordering problem. The question remains whether this result of an ideal, necessarily simple example is transferable to much more complex practical situations.

If, however, the optimum solution is not reached, there are two possible reasons for it. The subject might think that further improvement is impossible, or he might be dissatisfied with the solution he has found, but is practically unable to improve on it. The first reason can be considered as valid, if the subject performing a practical or experimental task does not know what the real optimum solution is, and the experimenter instructing him in the task cannot or does not tell him. The second reason comes in when a subject tries a number of displacements to improve on a specific configuration, but all his efforts lead to a deterioration of the result. Although he is not content with the result, he decides to make no further attempts.

When the process is stopped by a subject because he thinks that further improvement is impossible, the tool has at least met the needs of our subject and can be qualified to his satisfaction. Whether this result is also satisfactory to the experimenter, depends on the discrepancy between various subjective end results or between the subjective result and the objective one that was put in previously. In the second case, when unsatisfactory situations cannot be improved, the tool is not quite what it should be, although even now it is not correct to attach direct blame to the program. This situation is more complex than it looks at first sight. As the practical problems that can possibly set in are difficult to describe and are best explained with the help of some actual examples, we will refrain from doing so at this moment and deal with it in the section on the results and discussion of the experiment.

8.4 A computerized and interactive reorderable matrix

The principle and the original implementation of the reorderable matrix were discussed in section two of Chapter 4. In addition to the originally manual method, Bertin (1983) already developed a primitive computerized version of the reorderable matrix. This idea was further elaborated at our faculty into a user-friendly interactive program. This program will be used in the next experiment in which the analytical possibilities of the reorderable matrix principle will be further investigated. In this program, a rectangular matrix is drawn on a computer screen. The original data to be analysed are read-in and displayed as circle symbols in the cells of the matrix, where the sizes of the circles are in proportion to the original numerical values. Specific cells can now be selected and moved to another position within the rectangular matrix. Rearranging takes place by first selecting a cell or group of cells and then moving this selection to a new position within the matrix. The "mouse" is used for selecting, as well as moving the cells. Immediate feedback is provided upon selection of a certain region, all dots located within the selected area become grey and a contour line is drawn around the activated area. When the selection is moved across the matrix, the contour line moves along with the movements of the mouse. As soon as the mouse button is released the matrix will be redrawn, showing the selected part in its new, correct position. As all elements within a specific column or row belong to a particular object or one of its features, this means that only complete rows and columns can be moved. When one cell is selected and moved in the vertical direction the complete row will move along with it; when the movement of the cell is in the horizontal direction, all elements of the same column will also be displaced in the same way. In addition to strictly horizontal or vertical displacement, a selection can also be moved along a sloping line, displacing the respective columns and rows along with it.

All actions undertaken are recorded and stored by the program. The purpose of this is twofold. For the user of the program it first creates the chance to "undo" performed actions. When an action or a sequence of actions does not produce the expected results, one can return to a configuration of the symbols in the matrix that was attained a number of "steps" earlier. This extended version of the undo function, which is a standard option of most programs that run on Apple Macintosh computers, can be regarded as part of the compatibility between program (or interface) and user. It partly offsets the fear of making mistakes and provides the possibility of having a look at the effect of a specific action without being tied to it. The second reason for recording the performance of subjects in the process of structuring the information, is to use such information in the analysis of the experimental data. In section 4.6 we stressed the compatibility between the provided system, program or tool and the human approach to the specific task that has to be performed. By studying the complete ordering process as reflected in the recorded "steps" or displacements, we expect to discover some of the strategic aspects used by subjects in their performance of the experimental task.

8.5 Experiment 1

Procedure

In this experiment a total of 8 experimentally determined sets of dichotomous data were used. As yet, we have used only dichotomous data because the interaction between the specific task aspect of clustering and the weighting of differently sized symbols would probably, at least at this moment, be too complex and therefore a disturbing factor in the experimental context. The solutions to the 8 sets that had to be structured were more or less ideal. The sets had to be structured according to both the previously defined principles of orderliness and association. The two tasks (clustering and linking the groups along a continuously increasing or decreasing function) could be performed simultaneously by starting to group dots in one corner of the matrix and chaining the second and subsequent groups to the first one. Although subjects were not informed on the "correct" original configuration of the data, they were told that all randomized sets could be turned into a more or less regular configuration.

After they had studied the extensive written instructions, subjects were allowed some 5 minutes to pick up the specific features of the program, such as selecting and moving a group of cells and reverting to a previous state of the matrix. In the instructions, the concepts of association and orderliness were elucidated and operationalised to the matrix context. In order to explain the practical applications of this somewhat abstract tool, the example in Figure 4.10 was embodied in the instruction. All subjects had to solve the same 8 matrices. Apart from the first matrix, which was the easiest one to solve, the matrices were counterbalanced between subjects. Subjects were instructed not to spend more than about 15 minutes on each of the matrices and were free to take a short pause whenever they liked. The complete experiment took about two to two and-a-half hours.

Stimulus material

The size of each matrix was 25 rows by 25 columns. A varying number of cells contained dot-like symbols, all other cells were empty. The number of dots in the different matrices ranged from 94 (about 15% of the total number of cells) to 156 (≈25%). The size of the matrix cells was 0.5 cm square. All dots were the same in size, with a diameter of about 0.4 cm. Cells were visually bordered by dotted grid lines. In the original matrices, the symbols were divided across the matrix according to a predefined pattern, giving a more or less coherent and grouped distribution along the diagonal (right-hand part of Figure 8.4a). Before the matrices were presented to the subjects in the experiment, the rows and columns were randomized (left-hand part of Figure 8.4a). The matrices were presented on an Apple Macintosh computer screen, with contour of the matrix, grid lines and symbols black against a white background.

Subjects

Six male students of the Faculty of Architecture and Building participated in this experiment. Subjects had acquired some experience with the computer system in an architectural design course, but were not familiar with the reorderable matrix program. The subjects were paid for participation.

Results and Discussion

Configurations at the end of the ordering process

The initial configurations and "correct" solutions of the 8 matrices are presented in Figure 8.4. A clear-cut division into two groups can be made.

The first group consists of matrices whose optimum configuration contains some distinct rectangular clusters. These clusters have only a small overlap (matrix e) or no overlap at all (matrices a and g). The other matrices contain clusters that show considerably more overlap. The groups in these remaining matrices are still rectangular in shape (matrices d and f) or they are more difficult to distinguish because of a gradual transition from one group to another. Because of this gradual transition, the elements of these matrices practically lie along one line. Only the width of the connecting line varies (matrix b versus c and h).

Thus, on the one hand we have introduced a shift of distinct rectangular clusters towards groups that are more difficult to delimit, while on the other hand, the position of the individual elements close in on the centre of a continuously increasing or decreasing line. In the first group, the aspect of ordering is emphasized, in the last group the accent is on the overall association between the two components. The three matrices of the first group were solved by all subjects (in 14 to 66 steps, with an average of 35.7 steps). Of the other matrices, b (range 35 to 126 steps; average 81.8), f (range 46 to 151; average 88.3) and h (range 39 to 105; range 77.8) were each completely solved by 2, subjects whereas c (range 38 to 142; average 70.3) and d (range 28 to 151; average 80.2) were not solved by anybody. There were some considerable differences in the solutions submitted by the various subjects.

Using Kendall's coefficient of concordance, W, we tested whether rank ordering of the 6 subjects on each of the matrices in the set were in agreement. The *number of steps* required by each subject to solve the matrix were ranked for each of the 3 easy matrices (solved by all), the *final subjective solutions* (difference between attained scores and optimum) were ranked, for each of the 5 unsolved matrices. For the three matrices that were solved, the coefficient of concordance was found to be significant at $p<0.05$ which means that rankings are not independent or that some subjects are systematically faster in solving these easy matrices while others are continually slower. Kendall's W was also found to be significant ($p<0.01$) for the 5 unsolved matrices. Relative performance of a subject (as compared to others) on one of the matrices is not independent of his relative performance on other matrices. Some subjects manage to continually reach better performances.

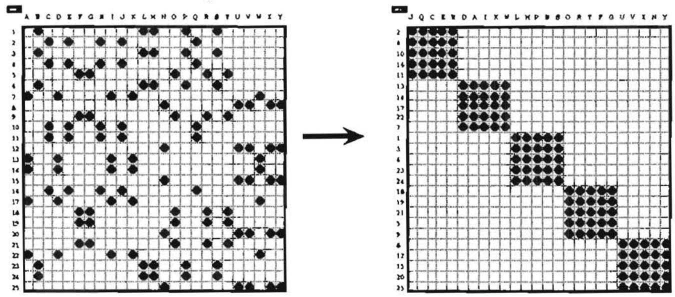


Figure 8.4 a

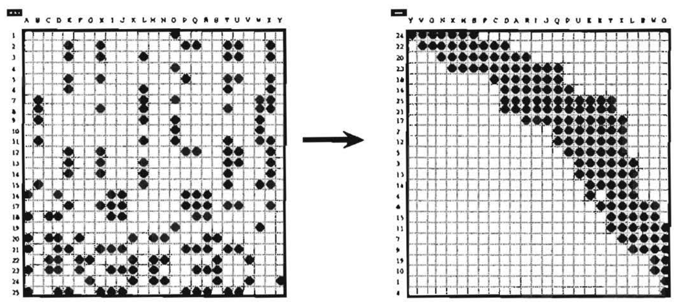


Figure 8.4 b

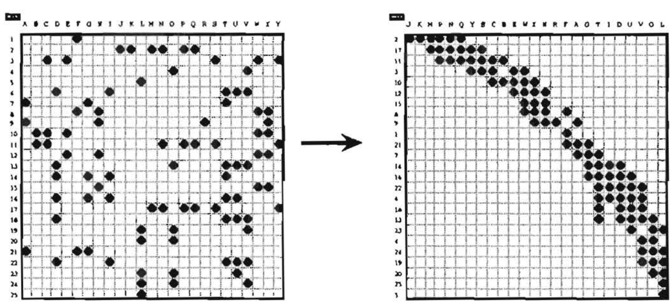


Figure 8.4 c

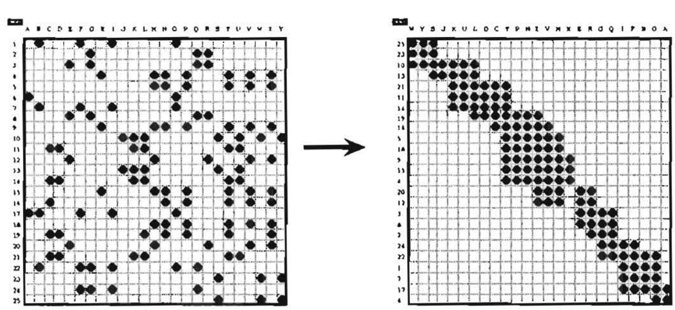


Figure 8.4 d

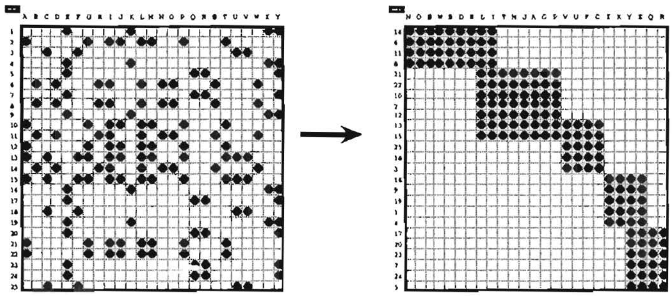


Figure 8.3 e

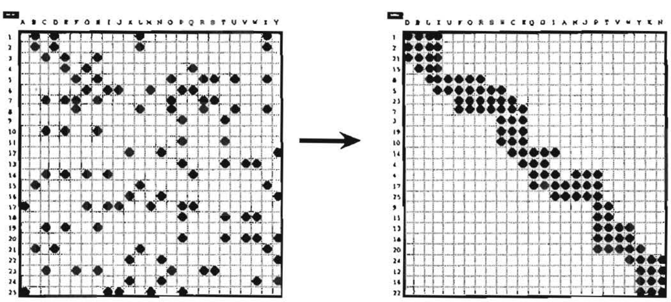


Figure 8.4 f

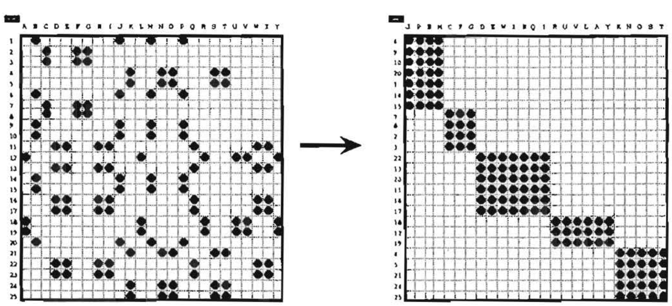


Figure 8.4 g

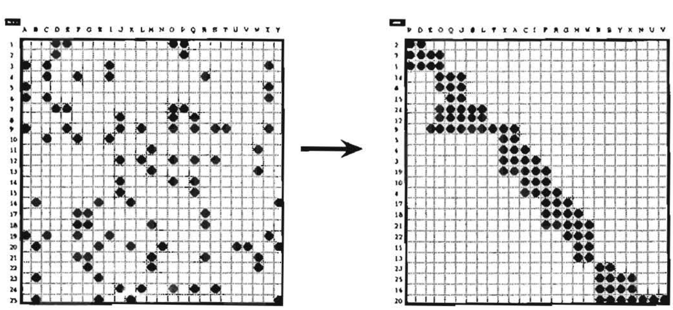


Figure 8.4 h

Figure 8.4 Original configurations as presented to the subjects (left-hand parts) and solutions to the matrices (right-hand part)

Some typical practical solutions are shown in Figure 8.5. Although only one example is shown for each of these typical final configurations, all could be observed several times, both with different subjects as well as with different matrices.

In Figure 8.5a (subject nr. 1 matrix f) the main axis of the clusters runs from the upper-left corner towards the lower-right one. The groups in the middle part, however, run from the lower left to the upper right. If the clusters in this middle part were simply turned over, the result would be much improved, and principally result in a higher score on the association value.

In Figure 8.5b (subject nr. 2 matrix b) the largest part of the symbols is grouped as a continuously decreasing function. One large group in the lower-left corner does not fit in with the rest. When the ordering process is studied, it can be seen that subjects have tried to link this "out-of-tune" cluster with the upper-left or lower-right part of the matrix, that is the parts with which the group has some elements in common on either the x or the y-axis. In both cases the situation deteriorates. The solution to this typical situation is that the complete collection of groups has to be moved up along the diagonal of the matrix. The groups located at the corners of the matrix are not correctly positioned and need to be moved towards the middle part of the matrix. When this is done correctly, the cluster that was out of place automatically fits in with the rest. Problems of this type can usually be solved by one single displacement.

In Figure 8.5c (subject nr.3 matrix f) the continuously decreasing function and diverse, coherent local clusters can be clearly distinguished. Some individual elements or small groups could not be joined with the rest as they apparently did not fit in. The solution to this problem is comparable to the previous one. It can be seen that the matrix can be divided into a small number of groups that show a slight overlap or none at all. Some simple interventions in which the relative positions of these groups are changed, normally result in a distribution that can easily be solved. The main point in the solution of this and the previous type of configuration is that subjects seem to be guided too much by local aspects of the matrix. They want to join the out-of-tune elements or groups directly to the rows or columns that have some elements in common. This is done by displacement of such element or group in either the strictly horizontal or vertical direction. The right solution is a change in the position of the diverse groups. For example, groups that are located at the end of the string in one of the corners of the matrix should be positioned more towards the middle.

The clustering task has been carried out fairly well in the matrix of Figure 8.5d (Subject nr. 4 matrix f). A final linking of the clusters apparently presented some more problems. Despite some effort by the subject in linking groups from the upper-left towards the lower-right corner, most of these groups are not correctly positioned. The solution to this type of configuration raises some more problems. This is mainly due to the relative position of the specific rows and columns within a cluster. In addition to a rearrangement of the groups, the mutual position of rows and columns within the cluster also has to be changed. Subjects have mainly considered the clustering aspect of the task and did not take the joining of these clusters into account until the end of the ordering process. Solving these clusters therefore might require quite a number of operations. In some of these departures from the optimum situation, one or more actions that lead away from the objective are required before the objective can be

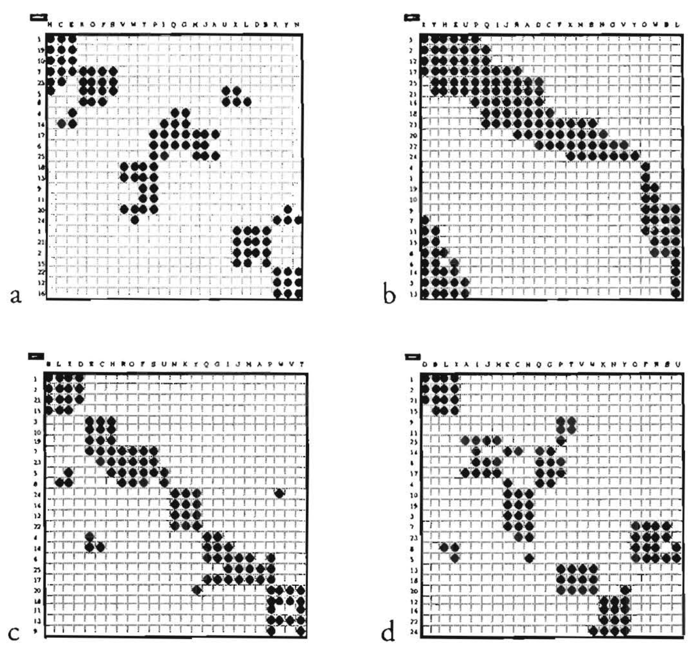
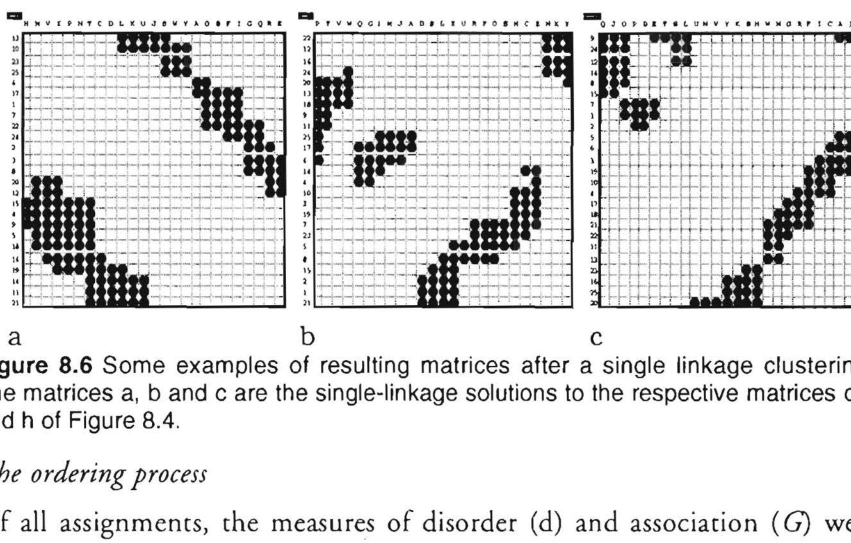


Figure 8.5 Some typical subjective solutions to the matrices. Matrices a, c and d are the solutions by different subjects to matrix "f" of Figure 8.4, matrix b is the solution by one of the subjects to matrix "b" in Figure 8.4

attained. This "means-end analysis" strategy can be observed in various problem-solving tasks (e.g. see Kohler, 1925, Chapters 1 and 2; Mayer, 1983; and The General Problem Solver of Newell and Simon, 1972).

Results of a statistical clustering analysis

In addition to the subjective solutions to the ordering process, some statistical clustering algorithms were tried, such as direct joining, k-means, single, median and complete linkage (Hartigan, 1975; Everitt, 1980). Of these methods, the single-linkage algorithm, generally provided the most satisfactory results. This method tends to produce long, stringy clusters. Of the matrices used in the experiment, a, e and g were solved by this method (although the correctly discovered groups still had to be joined), the results in some of the other matrices are presented in Figure 8.6.



a b c

Figure 8.6 Some examples of resulting matrices after a single linkage clustering. The matrices a, b and c are the single-linkage solutions to the respective matrices d, f and h of Figure 8.4.

The ordering process

Of all assignments, the measures of disorder (d) and association (*G*) were calculated with every change in the arrangement of elements.

$$d = \sum_{j=1}^{k-1} \sum_{i=1}^{r} \left| n_{i,j} - n_{i,j+1} \right| + \sum_{i=1}^{r-1} \sum_{j=1}^{k} \left| n_{i,j} - n_{i+1,j} \right| \qquad [8.4]$$

In this equation, the distance between two adjoining cells, in the horizontal direction between $n_{i,j}$ and $n_{i,j+1}$; in the vertical direction between $n_{i,j}$ and $n_{i+1,j}$ is defined as $|n_{i,j} - n_{i,j+1}|$, respectively $|n_{i,j} - n_{i+1,j}|$ where ($n_{i,j}$, $n_{i,j+1}$ and $n_{i+1,j}$ are the frequencies or the sizes of the elements in the respective cells).

The first component of the right-hand part expresses the total distance between columns, the second indicates the total distance between rows. The variable **r** equals the number of rows and **k** is the number of columns (see also equation 7.4).

$$G = \frac{\sum_{i<r} \sum_{j<k} n_{ij}\ n_{rk} - \sum_{i<r} \sum_{j>k} n_{ij}\ n_{rk}}{\sum_{i<r} \sum_{j<k} n_{ij}\ n_{rk} + \sum_{i<r} \sum_{j>k} n_{ij}\ n_{rk}} \qquad [8.5]$$

In this equation ($\sum_{i<r} \sum_{j<k} n_{ij}\ n_{rk}$) is the number of agreements or concordant pairs and ($\sum_{i<r} \sum_{j>k} n_{ij}\ n_{rk}$) is the number of disagreements or discordant pairs (see section 8.3.2).

The resulting values of one of the easier matrices are given in Figure 8.7. Curves of the results of three subjects are shown. Even though all three subjects reached the minimum orderliness value, or the optimum solution to this problem, there are considerable differences in effectiveness. Subject nr. 6 reached the optimum configuration after 23 steps. The curve of his performance, represented by white triangles, shows a constant increase. Subject nr. 5 took almost twice as many steps, his performance showing a curve that is less steep over the whole range, even sometimes flat in parts. The curve of subject nr. 3 lies midway between the previous two. His effectiveness is comparable to that of subject nr. 6 in the first part of the process but, the two curves gradually separate in the second part. The black symbols in this figure represent the association value in each of the configurations. Some interesting differences can also be noted here. The association curve of subject nr.6 shows a constant increase, which means that this subject is constantly improving the relative position of the clusters. The performance curve of subject nr. 5 is more irregular. During the first part of the process, the association value remains more or less at the level of the starting position; only after some 20 steps does the value show a gradual increase. Subject nr.3 scarcely pays any attention to the association value until the end of the ordering process. Some irregular improvements and deteriorations can be noted in the first part, the last 5 steps showing a rapid increase from the starting level to optimum performance. Obviously, this subject first arranged the individual elements in clusters before linking these clusters to a string.

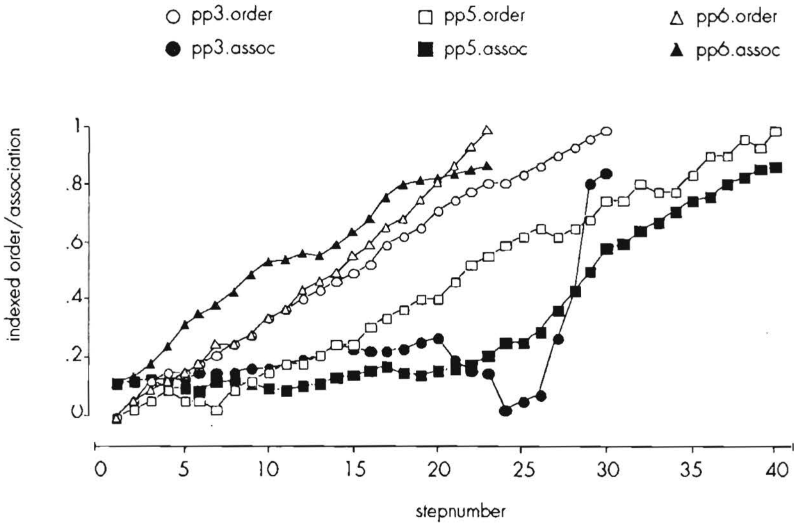


Figure 8.7 Values of ordering and association at each step in the structuring of a matrix. The results of three subjects are shown.

The same phenomena can be noted in Figure 8.8. This figure represents one of the more difficult matrices. The performance on the aspect of order by subject nr. 3 shows a very steep incline at first, but levels off after some 30 operations and very soon afterwards the subject stops the process. This subject again does not start linking the clusters until after some 35 steps. At the end of the process the optimum has not yet been reached. The resulting configuration in his case is the same as that presented in Figure 8.5c. As the matrix still contains some outstanding elements, the relative positions of the clusters are not completely correct. Subject nr. 5 is more persistent in trying to attain an optimum configuration. Neither of his two performance curves reaches the maximum value and the last 40 steps hardly bring any improvement to the results. After this subject corrected a serious drop in the association value he decided to stop the ordering process. Subject nr. 6 is again the most successful; he reaches the highest order and association values in the fewest number of steps.

The results of other matrices are comparable to those in Figures 8.7 and 8.8 (and are therefore not shown here). Performance by the other three subjects (numbers. 1, 2 and 4) stood more or less midway between that of subjects 6 and 5. Relative effectiveness (increase in order and association values divided by the number of steps in the analysis) of the diverse subjects was very consistent across the experimental conditions (matrices). The results of subject nr. 3, for example, were consistently better than those of other subjects. His relative effectiveness, which has been demonstrated in the resulting configurations after ordering, and in the curves that show the progress of the two important ordering aspects in the process itself, can be supported further by a number of typical, deviating actions in the ordering process. This subject often moved selections diagonally as opposed to the mainly strictly horizontal or vertical displacements of a selection by other subjects. A second remarkable action could be seen when one deviating row or column split an otherwise coherent group in two. Instead of selecting one of the parts of the group and moving it towards the other (as most subjects normally did), this subject just selected the deviating part and moved it towards a similar row or column. In this way the planned cluster was also obtained and at the same time the deviating part was positioned at the right location. Apparently this subject could correctly survey the results of his relatively complicated actions.

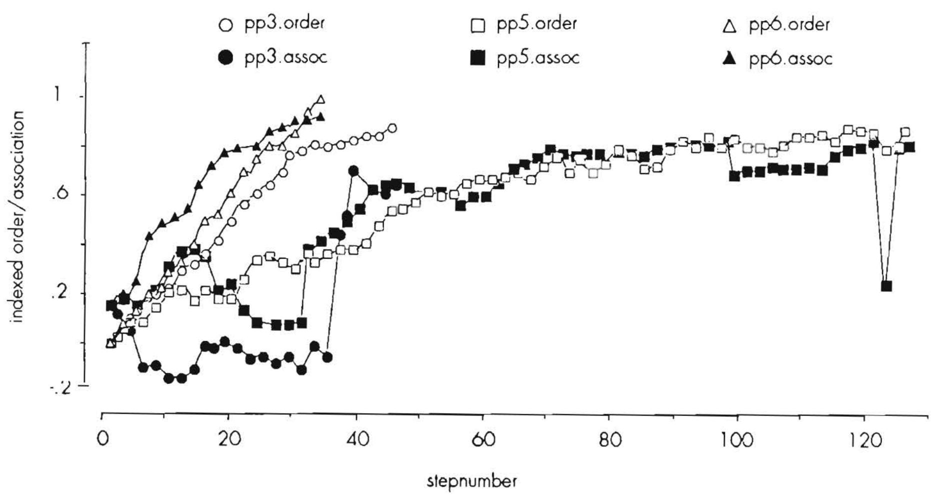


Figure 8.8 Association and ordering values after each change in the layout of elements in the matrix. Results of the ordering process of three subjects are shown.

8.6 Conclusion

Subjects are fairly capable of visually and interactively structuring the elements of a graphical matrix. The specific actions used for this purpose largely determine the effectiveness of the ordering process. Strategies used by diverse subjects proved to be quite consistent throughout the experiment. The process can be considerably increased, even maximised, when more complicated actions are used, such as moving selected areas diagonally. Since the resulting configurations of elements are, however more difficult to survey with these actions, there should be investigation into whether correct prediction of the effects of specific actions is feasible.

It is recommended that the ordering of elements in groups and linking the groups should be carried out from the very beginning of the process. Postponing the second aspect to a later stage in the process can require too many actions if the relative positions of rows and columns within a group happen to be incorrect.

When restricted, but more or less optimal sets of data are used, the results of some simple clustering algorithms can prove helpful in organising the data. The applicability of these algorithms to specific practical situations should be further investigated.

Chapter 9 Conclusions and Implications

9.1 Introduction

In this final evaluation we first discuss the method of the reorderable matrix and its implications in the presentation and analysis of multivariate data. This necessarily consists mainly of a summarized enumeration of the discussion sections of Chapters 3 to 8 inclusive. In the second part, attention will be claimed by computerization of the tool that enables graphical elements in a rectangular matrix to be reorganized. The practical implications of method and tool in the decision-making process are discussed in the third part.

9.2 Evaluation of the method

9.2.1 Aspects of presentation at the level of individual elements

Discrimination

In order to be able to visually detect a difference between objects, the minimum size of such difference has to be a specific fraction of the size of one of the elements. Psychophysical research on this problem shows that the percentage of correct detections of a difference gradually changes as a function of the actual relative difference (Weber Law). This rule appeared to be perfectly applicable to graphical elements, both when two of them were presented pairwise and when a larger number of these elements were displayed in a matrix.

Results of the experiments of Chapter 5 showed that the size ratio of 1.12, proposed by Bertin, has not been derived directly from the Weber ratio. A much smaller ratio suffices for a visual discrimination between two symbols of elementary shape. From a size ratio of about 1.10 upwards, the performance of a simple detection task (a pair of elements) and a complex detection task (larger group of symbols in a matrix) run almost parallel, provided that both functions are corrected for chance. At this specific ratio, a corrected performance of about 90% correct can be noted for circle symbols. These results imply that graphical symbols are perfectly suitable for a discrimination task, provided that their differences are not too small.

When, however, original numerical values cover a broad range and at the same time very small differences need to be detectable, a direct graphical translation must be discouraged.

The translation of numbers into a graphical format offers two major improvements. The speed at which graphical symbols can be visually compared is much higher than visual comparison of numbers, since a processing of the graphical information at a lower level suffices. In addition, large graphical symbols are much more conspicuous than large numbers.

Graphical symbols that show a unidimensional growth (e.g. bars), require smaller differences in areal size than symbols that grow simultaneously in two directions (e.g. circles and squares). Bars are therefore to be preferred to circles or squares in discrimination tasks.

Estimation

The usefulness of graphical symbols in detection tasks can be pursued when retrieval of original values is at stake. The size of graphical symbols can be accurately estimated, provided that the matrix is accompanied by a legend which covers the complete range of symbol sizes. Power function exponents, calculated to describe the estimation of symbol sizes, were very close to an optimum value of 1.0. This value applied to simply shaped graphical symbols, such as bars, dots and squares. Bertin's value of 1.12 therefore, seems not to be derived from the exponent of the power law. Two things should be noted in addition.

- Although the size of graphical elements can be determined fairly accurately, slight overratings and underestimations often occur. Unless a very precise determination of represented values is essential, graphical symbols are quite a practical alternative to numbers. When precise determination is required, graphical symbols could be combined with their numerical counterparts in a standard graphical display with the exact value they represent "repayable at call". This notion should be considered against the observation on the minimum required difference that was previously made.
- Subjects tend to use round numbers (e.g. multiples of five or ten) when estimating the size of graphical symbols. This tendency calls for a categorization of original values and a matching range-graded method in their graphical translation.

Separation of matrix cells by grid lines had an ambiguous and detrimental effect on the accuracy of estimation. As far as judgment of size is concerned, the use of grid lines should be rejected.

9.2.2 Aspects of presentation at the intermediate and overall level of the matrix

Intermediate

A model correctly predicting subjectively perceived groups of graphical elements displayed on thematical maps, appeared to be inadequate for the prediction of similar groups in a matrix. However, most of the symbol pairs that were predicted, but not perceived as clustered, showed some specific characteristics that were not met by the perceived pairs. This offers the possibility of elaborating the original prediction model to one that is also useful in a matrix context. The specific characteristics mentioned, refer to the number of elements in a group and the contiguity of its outline. Incorporation of these variables implies that the original starting point of the stimulus pair (based on the principle of Delaunay neighbours) has to be abandoned. Larger groups than a simple pair of nearest neighbours in the sampling network have to be taken into account.

Overall

There is a high level of agreement between subjects when it comes to rating the overall structure of graphical circle elements in a matrix. When the best alternative layout has to be selected this allows a very fast and accurate decision. In addition, subjects seem to be able to give rather a precise rating of differences and ratios between pairwise presented symbol configurations.

A fairly correct prediction of perceived clusters and an even more accurate prediction of subjectively estimated structures make the graphical matrix very useful as a means of condensing the communication of large amounts of data. A close resemblance between intended messages and the reader's impression of spatial information are a first gain and allow fruitful discussion when making decisions.

Validation

Subjective impressions of the coherence of elements in groups and the overall structure of elements in a matrix showed a highly significant

correlation with objectively defined measures of order (groups) and association (overall structure).
The, relatively simple, objective measure of order was based on the geometric distance between elements (city-block metric distance function), standard statistical measures of association, just as the coefficient of concordance and the gamma statistic G were used as the yardstick for rating the subjective impression of overall structure. As both objective measures showed a direct functional relationship to the practical purposes of the reorderable matrix, they seem to provide valid criteria in the analysis of authentic multivariate data. Thus, a correlation between subjective impressions and objectively defined statistical measures validate the use of the reorderable matrix as an accurate communicative device in the presentation of information at the intermediate and overall levels of a matrix.

9.2.3 Aspects of analysis

The reorderable matrix seems to be one of the very few graphical methods that truly warrants its claim to being an *analytical* tool. The utility of most of the methods discussed in Chapter 3 is, on the contrary, restricted to one of graphic *communication.* The reorderable matrix first transcribes all the data from the original numerical table and second, its interactive nature allows of an investigation into all pertinent questions and the involvement of extrinsic information. Taken alone, these two prerequisites are not sufficient to enable one to appreciate the reorderable matrix as practically useful and attainable. At the higher level of information processing especially (the level that Bertin calls "a simplification of the two components of the table.") we have to exercise the greatest caution. Simplification is admittedly necessary, but sufficient is still a far cry from genuine simplification.

When the interactive aspect is introduced, the potentialities of the reorderable matrix are extended; mere visual detection and recognition tasks are supplemented by problem-solving tasks. Heuristic strategies have to be developed to "solve" the configurations of graphic elements. The experiment described in Chapter 8 revealed considerable differences in the heuristic strategies being followed.

The analytical possibilities of the reorderable matrix seem to be promising. At least one of the subjects appeared to follow an effective (correct) heuristic strategy and solved most of the matrices. The matrices

that were not solved, either by this subject or others, could be roughly assigned to a few categories, with all categories showing some typical departure from the optimum solution. For each of these categories it proved to be possible to indicate the specific actions to be performed in order to reach a more or less optimum solution after all. In addition, some aspects in the strategies of the more effective and more ineffective subjects were demonstrably related to the (in)ability to solve the structuring problems. Use of more complicated actions, shifting attention from local to global aspects and working at two tasks at a time (clustering elements and joining clusters) were characteristics that could improve the effectiveness of the task performance when used correctly.
It remains to be investigated as to how far subjects are able to survey or predict the results of performed or intended actions and to what extent this insight can be acquired. The relative contribution of visual and interpretative aspects in different phases of the analysis also have to be further examined. If the purely visual parts could be accurately operationalised, it would be possible to support them by specific statistical algorithms.

9.3 Evaluation of the tool

The practical application of the computerized version of the reorderable matrix was restricted to one single experiment. Nonetheless, it was evidently shown that subjects generally used only the most elementary, simple options that the program provided. Whereas some of the incorrect or ineffective actions should be put down to inexperience, others seem to be more related to spatial insight and the possibility of predicting the outcome of intended actions. Further research on subjects' progress in performance, acquired through experience or through instructions, is needed before preparations are made to extend the program to include sophisticated but untried options.

All things considered, the reorderable matrix appears to be a useful tool as it is based on starting points experimentally found to be correct. As the data in most experimental conditions, however, were restricted to experimentally defined sets and most participants in the experiments can, as students of Architecture, be regarded as more or less visually oriented, the extent of the tool's suitability in realistic situations has to be further investigated.

9.4 The role of the reorderable matrix in decision making

The method of the reorderable matrix should not, in the first place, be considered as an alternative to statistical methods; the respective approaches differ fundamentally and should be used within their own contexts. The reorderable matrix is principally suited to less stringent inquiries, where ascertainment of a statistical significance is not the primary purpose. Exploratory study of data where interest is directed towards various levels of the information (individual elements, groups, overall association) can be very properly performed by this graphical, interactive approach as well as research where extrinsic information on the displayed objects and their characteristics is, or can be an important factor. The matrix can also be used in combination with statistical techniques. Preliminary work, such as organization and structuring of the data (based on visual characteristics) and generation of hypotheses (based on extrinsic information) can be reserved to visual interactive analysis, whereas eventual tests of accurately defined hypotheses could be done by applying the appropriate algorithms.

In addition to research where the statistical and graphical interactive approach can be used side by side, the last-mentioned approach has some specific characteristics that certify its usefulness as a stand-alone tool in decision making. As all individual information items are available throughout the analysis, pros and cons of a nominated alternative can be directly and easily weighed against one another. This opens up the possibility of discussing specific subjective choices and decisions and allows of an investigation of considerations that determined these choices and decisions.

References

Adams, W.
An Ergonomical Approach to Bertin's Method of Graphically Displaying Multivariate Data. Information Design Journal, 1988, vol. 5 no 3, pp. 1-11.

Adams, W. & Daru, R.
Facility Planning Nieuwe Stijl. VCA Mededelingen, februari 1990, 19-42.

Agresti, A.
Analysis of Ordinal Categorical Data. John Wiley & Sons, 1984.

Albers, L.H.
Het Gewichtloze gewogen: Cultuurhistorische Betekenis van Landgoederen Geevalueerd met behulp van Multicriteria Analyse. Delftse Universitaire Pers, 1987.

Anderson, E.
A semi-graphical Method for the Analysis of Complex Problems. Technometrics, 1960, 2, 387-391.

Andrews, D.F.
Plots of High-Dimensional Data. Biometrics, 1972, vol. 28, 125 -136.

Attneave, F.
Some Informational Aspects of Visual Perception. Psychological Review, 1954, 61, 183-193.

Baird, J.C.
Area and Distance Estimation of Single and Multiple Stimuli. Vision Research, 1965, 5, 661-678. .

Baird, J.C.
Psychophysical Analysis of Visual Space. Pergamon Press, New York, 1970.

Bertin, J.
Sémiology Graphique. Editions Gauthier-Villars, Paris, 1967.

Bertin, J.
La Graphique et le Traitement Graphique de l'Information. Flammarion, Paris, 1977.

Bertin, J.
Semiology of Graphics. University of Wisconsin Press, Madison, WI, 1983.

Bertin, J.
Grafische Semiologie. W. de Gruyter, Berlin. New York, 1974.

Bertin, J.
Graphics and Graphic Information Processing W. de Gruyter, Berlin. New York, 1981.

Bruckner, L.A.
On Chernoff Faces. In: P.C.C. Wang (ed.), Graphical Representation of Multivariate Data. Academic Press, 1978. pp. 93-121.

Card, S., Moran, T.P. & Newell, A.
The psychology of Human-Computer Interaction. Lawrence Erlbaum Associates, Hillsdale, N.J., 1983.

Chambers, J.M.
Computational Methods for Data Analysis. Wiley, New York, 1977.

Chambers, J.M., Cleveland, W.S., Kleiner, B. & Tukey, P.A.
Graphical Methods for Data Analysis. Wadsworth, Belmont, CA., 1983.

Chambers, J.M. & Kleiner, B.
Graphical Techniques for Multivariate Data and for Clustering. In: P.R. Krisnaiah (ed.), Handbook of Statistics. Volume II. North-Holland, New-York, 1982.

Chang, K.T.
Data Differentiation and Cartographic Symbolization. The Canadian Cartographer, 1976, 13, 60-68.

Chang, K.T.
Visual Estimation of Graduated Circles. The Canadian Cartographer, 1977, 14, 130-138.

Chang, K.T.
Measurement Scales in Cartography. The American Cartographer, 1978, vol. 5, 57-64.

Chang, K.T.
Circle Size Judgment and Map Design. The American Cartographer, 1980, vol. 7, 2, 155-162.

Chernoff, H.
The Use of Faces to Present Points in k-dimensional Space Graphically. Journal of American Statistical Association, 1973, vol. 68, 361-368.

Chernoff, H.
Graphical Representations as a Discipline. In: P.C.C. Wang (ed.), Graphical Representation of Multivariate Data. Academic Press, 1978.

Chernoff, H. & Rizvi, H.M.
Effect on Classification Error of Random Permutations of Features in Representing Multivariate Data by Faces. Journal of the American Statistical Association, 1975, 70, 548-554.

Cleveland, W.S., Harris, C.S., & McGill, R.
Judgments of Circle Size on Statistical Maps. Journal of the American Statistical Association, 1982, vol. 77, 379, 541-547.

Cleveland, W.S., Harris, C.S., & McGill, R.
Experiments on Quantitative Judgments of Graphs and Maps. The Bell System Technical Journal, 1983, vol. 62, 6, 1659-1674.

Cohen, J. & Cohen, P.
Applied Multiple Regression / Correlation Analysis for the Behavioral Sciences. Lawrence Erlbaum, Hillsdale, N.J., 1983.

Cox, C.W.
Anchor Effects and the Estimation of Graduated Circles and Squares. The American Cartographer, 1976, vol. 3, 1, 65-74.

Cox, D.R. & Lauh, E.
A Note on the Graphical Analysis of Multidimensional Contingency Tables. Technometrics, 1967, 9, 481-488.

Crawford, P.V.
Perception of Grey-tone Symbols. Annals of the Association of American Geographers, 1971, 61, 4, 721-735.

Crawford, P.V.
The Perception of Graduated Squares as Cartographic Symbols. The Cartographic Journal, 1973, 10, 88.

Cross, D.V.
Sequential Dependencies and Regression in Psychophysical Judgments. Perception & Psychophysics, 1973, 14, 547-552.

D'Amato, M.R.
Experimental Psychology: Methodology, Psychophysics and Learning. McGraw - Hill, 1970.

Daru, R.
Interactive Graphic Heuristics. Paper presented at CAPE' 89, Tokyo, 1989.

Daru, R.
Evalueren met Grafische Middelen (in Dutch). Beleidsanalyse, 1983, 3, pp. 5-14.

Daru, R., & van Gils, A.
Decision Graphics: Criteria for Development and Evaluation. In: K. Bø, L. Estensen, P. Falster & E. Warman (eds.), Computer Applications in Production and Engineering. Elseviers Science Publishers, Amsterdam, 1987. pp. 913-922.

Dobson, M.W.
Refining Legend Values for Proportional Circle Maps. The Canadian Cartographer, 1974, vol. 11, 1, 45-53.

Dobson, M.W.
Eye Movement Parameters and Map Reading. The American Cartographer, 1977, vol. 4, 1, 39-58.

Dobson, M.W.
The Influence of Map Information on Fixation Localization. The American Cartographer, 1979, 6, 51-65.

Dobson, M.W.
Visual Information Processing During Cartographic Communication. The Cartographic Journal 1979, vol. 16, 1, p. 14-20.

Dobson, M.W.
The Influence of the Amount of Graphic Information on Visual Matching. The Cartographic Journal, 1980, 17, 1, p. 26-32.

Dobson, M.W.
Benchmarking the Perceptual Mechanism for Map-reading Tasks. Cartographica, 1980, vol. 17, 1, 88-100.

Ekman, G., Lindman, R. & William-Olsson, W.
A Psychophysical Study of Cartographic Symbols. Perceptual and Motor Skills, 1961, vol. 13, 355-368.

Ekman, G. & Junge, K.
Psychophysical Relations in Visual Perception of Length, Area and Volume. Scandinavian Journal of Psychology, 1961, vol. 2, 1-10.

Engel, F.L.
Visual Conspicuity as an External Determinant of Eye Movements and Selective Attention. Thesis, Eindhoven, The Netherlands, 1976.

Engen, T.
Psychophysics. Discrimination and Detection. In: J.W. Kling & L.A. Riggs (eds.), Experimental Psychology. Holt, Rinehart & Winston, N.Y., 1971 (3rd ed.)

Everitt, B.S.
Graphical Techniques for Multivariate Data. Heinemann Education Books, 1978.

Everitt, B.S.
Cluster Analysis. Halsted Press, 1980 (2nd edition).

Falzon, P.
From Surface Form to the Structure of the Interface Studies of Human Computer Interaction at INRIA., in: G. Salvendy (ed.), Human Computer Interaction. Elsevier Science Publishers, Amsterdam, 1984.

Fienberg, S.E.
Preliminary Graphical Analysis and Quasi-Independence for Two-Way Contingency Tables. Applied Statistics, 1969, 18, pp. 153-168.

Fienberg, S.E.
The Analysis of Cross-Classified Categorical Data. MIT Press, Cambridge, Mass., 1977.

Fienberg, S.E.
Graphical Methods in Statistics. The American Statistician, 1979, 33, pp. 165-178.

Fischer, H. T.
Mapping Information. Abt Books, Cambridge, Ma., 1982.

Flannery, J.J.
The Graduated Circle: A Description, Analysis and Evaluation of A Quantitative Map Symbol. Doctoral Dissertation, University of Wisconsin Press, 1956.

Flannery, J.J.
The Relative Effectiveness of Some Common Graduated Point Symbols in the Presentation of Quantitative Data. Canadian Cartographer, 1971, vol. 8, 96-109.

Flury, B. & Riedwyl, H.
Graphical Representation of Multivariate Data by Means of Asymmetrical Faces. Journal of the American Statistical Association, 1981, vol. 76, p. 757-765.

Funkhouser, H.G.
Historical Development of the Graphical Representation of Statistical Data. Osiris, 1937, 3, pp. 269 -404.

Garner, W.R.
The Processing of Information and Structure. Lawrence Erlbaum, 1974.

Gick, M.L. & Holyoak, K.
Problem Soving and Creativity. In: A. L. Glass, K.J. Holyoak & J.L. Santo, Cognition. Addison - Wesley, London, 1979.

Gilmartin, P.P.
Influences of Map Context on Circle Perception. Annals, Association of American Geographers, 1981, 71.

Gilmartin, P.P.
The Interface of Cognitive and Psychophysical Research in Cartography. Cartographica, vol. 18, 3, 1981, 9-20.

Goodman, L.A. & Kruskal, W.H.
Measures ofAssociation for Cross Classifications. Journal of the American Statistical Association, 1954, vol. 49, pp. 732-764.

Goodman, L.A. & Kruskal, W.H.
Measures ofAssociation for Cross Classifications. Springer Verlag, N.Y., 1979.

Green, P.E., Carmone, F.J. & Smith, S.M.
Multidimensional Scaling: Concepts and Applications. Allyn & Bacon, Boston, 1989.

Guilford, J.P.
Psychometric Methods. McGraw - Hill, 1954.

Haber, R.N. & Hershenson, M.
ThePsychology of Visual Perception. Holt, Rinehart & Winston, 1980 (2nd edition).

Hartigan, J.A.
Clustering Algorithms. Wiley, New York,1975.

Höffler, H.
Problem-Darstellung und Problem-Lösung in der Bauplanung: Entwicklung der Voraussetzungen eines Handbuches für Formalisierte Modell- und Darstellungstechniken. IGMA Dissertationen 3. Karl Krämer Verlag Stuttgart/Bern 1972.

Huff, D.L. & Black, W.
A Multivariate Graphic Display for Regional Analysis. In: P.C.C. Wang (ed.), Graphical Representation of Multivariate Data. Academic Press, 1978, p. 199-218.

Israëls, A.Z., Keller, W.J.
Multicriteria analyse voor ordinale data, Kwantitatieve methoden, 1986, 21, 7, pp. 49-74.

Jacob, R.J.K., Egeth, H.E. & Bevan, W.
The Face as a Data Display. Human Factors, 1976, 18, pp. 189-200.

Jacob, R.J.K.
Facial Representation of Multivariate Data. In: P.C.C. Wang (ed.), Graphical Representation of Multivariate Data. Academic Press, 1978, pp.143-168.

Jenks, G.F.
The Evaluation and Prediction of Visual Clustering in Maps Symbolized with Proportional Circles. In: J.C. Davis & M.J. McCullagh (eds.), Display and Analysis of Spatial Data. John Wiley, London, 1975.

Kleiner, B. & Hartigan, J.A.
Representing Points in Many Dimensions by Trees and Castles. Journal of the American Statistical Association, 1981, vol. 76, 260-276.

Kosslyn, S.M.
Graphics and Human Information Processing: A Review of Five Books. Journal of the American Statistical Association, 1985, 80, 391, pp. 499-512.

Kohler, W.
The Mentality of Apes. Routledge, Kegan Paul, London, 1925.

Leeuwenberg, E.L.J.
A Perceptual Coding Language for Visual and Auditory Patterns. American Journal of Psychology, 1971, 84, pp. 307-349.

Liggett, R.S.
Optimal Spatial Arrangement as a Quadratic Assignment Problem. Design Optimization, 1985, pp. 1-40.

Liggett, R.S. & Mitchell, W.J.
Optimal Space Planning in Practice. Computer Aided Design, 1981, 13, 5, pp. 277-288.

LLoyd, R.E.
A Look at Images. Annals of the Association of American Geographers, 1982, 72, 4, pp. 532-548.

Lloyd, R.E. & Steinke, T.R.
The Decision-Making Process For Judging The Similarity of Choropleth Maps. The American Cartographer, 1976, 3, 2, pp. 177-184.

Lloyd, R.E. & Steinke, T.R.
Visual and Statistical Comparison of Choropleth Maps. Annals of the Association of American Geographers, 1977, 67, 3, pp. 429-436.

Loether, H.J. & Tavish, D.G.
Descriptive Statistics for Sociologists. Allyn & Bacon, Boston, 1974.

Macdonald-Ross, M.
How Numbers are Shown: A Review of Research on the Presentation of Quantitative Data in Texts. Audio-Visual Comm. Rev., 1977, 25, 4, pp. 359-409.

MacMillan, N.A., Moschetto, C.F., Bialostozky, F.M. & Engel, L.
Size Judgment: The presence of a standard increases the exponent of the power law. Perception & Psychophysics, 1974, vol. 16, 2, 340-346.

Marks, L.E. & Cain, W.S.
Perception of Intervals and Magnitudes for Three Prothetic Continua. Journal of Experimental Psychology, 1972, 94, 6-17.

Mashour, M. & Hosman, J.
"On The 'New Psychophysical Law': A Validation Study". Perception and Psychophysics, 1968, 3, p. 367-375.

Massaro, D.W.
Experimental Psychology and Information Processing. Rand McNally, Chicago, 1975.

Massaro, D.W. & Anderson, N.H.
Judgmental Model of the Ebbinghaus Illusion. Journal of Experimental Psychology, 1971, 89, 147-151.

Mayer, R.E.
Thinking, Problem Solving, Cognition. Freeman and company, New York, 1983.

Meihoefer, H.J.
The Utility of the Circle as an Effective Cartographic Symbol. Canadian Cartographer, 1969, vol. 6, 105-117.

Meihoefer, H.J.
The Visual Perception of the Circle in Thematic Maps: Experimental Results. Canadian Cartographer, 1973, vol. 10, 63-84.

Mezzich, J.E. & Worthington, D.R.L.
A Comparison of Graphical Representations of Multi-Dimensional Psychiatric Diagnostic Data. In: P.C.C. Wang (ed.), Graphical Representation of Multivariate Data. Academic Press, 1978, pp. 123-141.

Muller, J.C.
Associations in Choropleth Map Comparison. Annals, Association of American Cartographers, 1975, vol. 65, no. 3, pp. 403-413.

Muller, J.C.
Bertins Theory of Graphics/ A Challenge to North American Thematic Cartography. Cartographica, 1981, 18, 3, p. 1-8.

Newell, A. & Simon, H.A.
Human Problem Solving, Prentice Hall, Englewood Cliffs, N.J., 1972.

Nijkamp, P.
Multidimensional Spatial Data and Decision Analysis. John Wiley, 1979.

Oeffelen, M.P.J. van
The Perception and Processing of Visual Numerosity. Phd Dissertation. Krips Repro, Meppel, 1984.

Olson, J.M.
Autocorrelation and Visual Map Complexity. Annals, Association of American Cartographers, 1975a, vol. 65, no. 2, pp. 189-204.

Olson, J.M.
Experience and the Improvement of Cartographic Communication. The Cartographic Journal, 1975b, 12, 94-108.

Olson, J.M.
Rescaling Dot Maps for Pattern Enhancement. International Yearbook of Cartography, 1977, 17, pp. 125-136.

Poulton, E.C.
The New Psychophysics: Six Models for Magnitude Estimation. Psychological Bulletin, 1968, vol. 69, no. 1, pp. 1-19.

Ramsay, J.O.
Intra- and Interindividual Variation in the Power Law Exponent for Area Summation. Perception and Psychophysics, 1979, 26, pp. 495-500.

Robinson, A.H. & Sale, R.S.
Elements of Cartography. John Wiley, NY, 1969 (3rd ed.).

Schmid, C.F. & Schmid, S.E.
Handbook of Graphic Presentation. John Wiley, NY, 1979 (2nd edition).

Schmid, C.F.
Statistical Graphics. John Wiley, NY, 1983.

Shortridge, B.G. & Welch, R.B.
Are We Asking the Right Questions? The American Cartographer, 1980, 7, pp. 19-24.

Siegel, J.H., Goldwyn, R.M. & Friedman, H.P.
Pattern and Process of the Evolution of Human Septic Shock. Surgery, 1971, 70, pp. 232-245.

Siegel, S. & Castellan, N.J.
Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill, 1988 (2nd edition).

Sjöberg, L.
Three Models for the Analysis of Subjective Ratios. Scandinavian Journal of Psychology, 1971, 12, pp. 217-240.

Slocum, T.A. & Gilmartin, P.P.
Cartographic Analysis of Proportional Circle Maps Using Delaunay Triangles. The Canadian Cartographer, 1979, 16, pp. 133-144.

Slocum, T.A.
Predicting Visual Clusters on Graduated Circle Maps. The American Cartographer, 1983, 10, 1, pp. 59-72.

Stevens, S.S.
On The Psychophysical Law. The Psychological Review, 1957, vol. 64, 3, 153-181.

Stevens, S.S. & Greenbaum, H.B.
Regression Effect in Psychophysical Judgment. Perception & Psychophysics, 1966, 1, pp. 439-446.

Stevens, S.S. & Guirao, M.
Subjective Scaling of Length and Area and Matching of Length to Loudness and Brightness. Journal of Experimental Psychology, 1963, 66, pp. 177-186.

Stewart, J.Q.
Empirical Mathematical Rules Concerning the Distribution and Equilibrium of Population. Geographical Review, 1947, vol. 37, 461-485.

Storbeck, D.
Sozialökonomisches Gutachten über die Eignung der für die Universitätsgründung in Ostwestfalen vorgeschlagenen Standorte. In: Arbeitsberichte zur Planungsmethodik 1: Bewertungsprobleme in der Bauplanung. Institut zur Grundlagen der Modernen Architectur (IGMA), Universität Stuttgart. Karl Krämer Verlag Stuttgart/Bern 1969.

Teghtsoonian, M.
The Judgment of Size. American Journal of Psychology, 1965, 78, 392-402.

Teghtsoonian, M. & Teghtsoonian, R.
How Repeatable are Steven's Power Law Exponents for Individual Subjects. Perception & Psychophysics, 1971, 10, 147-149.

Theodorescu, D.
Le Chapiteau Ionique Grec: Essai Monographique. Genève: Librairie Droz, 1980.

Torgerson, W.S.
Theory and Methods of Scaling. John Wiley & Sons, 1958.

Tufte, E.R.
The Visual Display of Quantitative Information. Graphics Press, Cheshire, CT, 1983.

Tukey, J.W.
Exploratory Data Analysis. Addison-Wesley, Reading, Ma 1977.

van der Helm, P.
Accessibility and Simplicity of Visual Structures. PhD Thesis, Benda Nijmegen, 1988.

Verillo, R.T. & Graeff, C.K.
The Influence of Surface Complexity on Judgments of Area. Perception & Psychophysics, 1970, 7, 289-290.

Vroon, P.A.
Het verwarren van munten. In: W.A. Wagenaar, P.A. Vroon, & W.H. Janssen (eds.), Proeven op de som. Van Loghum Slaterus, Deventer, 1978, pp. 33-38.

Wainer, H. & Thissen, D.
Graphical Data Analysis. Annual Review of Psychology, 1981, vol. 32, 191-241.

Wang, P.C.C.
Graphical Representation of Multivariate Data. Academic Press, 1978.

Watanabe, S.
Pattern Recognition: Human and Mechanical. John Wiley, 1985.

Williams, L.G.
The Effect of Target specification on Objects Fixated during Visual Search. Acta Psychologica, 1967, 27, 355-360.

Author index

Summary

Graphics are generally recognized and appreciated as expedients in the presentation or communication of information (*graphic communication*). In addition to this function graphics are, in the present study, also asserted to be applicable as important tools in information analysis and decision making (*graphic processing*). Especially if the total of recorded information has to remain intact and clearly visible throughout these processes, for example in investigations of an explorative nature (generation of hypotheses as compared to testing them), or when weighing up different criteria is inconclusive (importance of safety and aesthetical aspects in comparison with cost of construction), graphics can be regarded as useful in discovering and clarifying underlying relationships. The French cartographer Jacques Bertin has developed a graphic tool which, he claims, meets the more traditional requirements of correct and surveyable presentation of the authentic information as well as its more innovative, simple and obvious interactive analysis. This *"reorderable matrix"* method, particularly suited to the analysis of multivariate data, is based on direct translation of individual numbers into simple graphical symbols, such as dots or squares. These symbols, whose sizes are in accordance with the original scores they represent, are restored in the cells of a data matrix or table and presented on a computer screen. The symbol in each matrix cell represents the score of a feature (in the rows of the table) on a specific object (in the columns). When the number of bedrooms (feature) in a specific house (object) equals 4, this value, or a symbol of corresponding size, is recorded in the relevant cell of the matrix. In the analysis of the data, the layout of the information can be rearranged by moving a selected part of the matrix, containing one or more of its rows and columns, to a new position. The matrix construction lends itself well to such, mathematically correct, *"permutations"*. As the matrix is redrawn on the computer screen after this movement, immediate visual feedback is given and the effectiveness of manipulations can be interactively assessed. In this method, complex cognitive processes in the analysis and interpretation of information are thus operationalised into tasks of visual discrimination, pattern recognition and interactive pattern formation, that are highly developed in people. Visually oriented subjects, such as architects, geographers in particular, should be able to profit from this graphical tool.

In the present study, the required aspects of presentation (correct reproduction of the original data, allowing an accurate interpretation of individual values, differences, similarities and associations) and analysis (leading to a simple and obvious reorganisation of the information) in the reorderable matrix were experimentally tested.

In Chapter 2 a possible domain of application is outlined with the help of two architectural examples.

The reorderable matrix method is, in chapter 3, compared to a number of graphical methods, all developed to present multivariate data. The various methods are examined for their potentialities in the presentation and, in particular, the analysis of multivariate information.

Chapter 4 is devoted to the graphical theory of Bertin and the different graphical constructions he proposes. Each of these instruments can be applied with, and is especially suitable for the presentation and solution of specific, sets of data, matrices or networks. One of these, the reorderable matrix, is further elaborated in the present study.

Experiments testing the different requirements of the reorderable matrix are described in Chapters 5 to 8 inclusive.

Criteria at the early-visual-processing levels of discrimination and recognition are discussed in Chapters 5 and 6. As regards the discrimination and interpretation of individual symbols of simple shape it was found that people are fairly accurate in these tasks of discriminating between, sorting and estimating the individual sizes of graphical symbols. Even though Bertin's recommendations regarding required size differences were found to be not directly derived from psychophysical laws, such as the Weber law or power law, for reasons of efficiency and ease, they proved to be practically useful.

Groups of elements that are similar in size and therefore recognized as forming a coherent cluster or pattern, could not be predicted very accurately by an initial model based on Gestalt principles, discussed in Chapter 7. Some obvious characteristics of the perceived patterns, however, offer prospects for a more efficient extension and refinement of this model.

Subjective ratings of the overall configuration of elements in a matrix, *the overall image*, were found to be very consistent both in time and as between subjects. The visually estimated degree of overall structure of a matrix thus proved to be a concept that is generally agreed upon. In addition, this subjectively noted order was found to be strongly correlated with objective differences, measured by elementary geometric-distance functions. Subjective ordering therefore corresponds to the objectively measurable degree of regularity and association of the data.

As to the interactive ordering of graphic matrix elements, subjects were found to be able to improve significantly on their original disarray. Even though there were considerable differences between subjects in the efficiency and speed of task execution, all subjects attained or came close to the experimentally defined initial layout.

All things considered, the reorderable matrix appears to be a useful tool as it is based on starting points experimentally found to be correct. As the data in most experimental conditions were restricted to experimentally defined sets, the extent of the tool's suitability in realistic situations, however, has to be further investigated.

Samenvatting

Grafieken worden algemeen erkend, gewaardeerd en veelvuldig gebruikt als hulpmiddel in de presentatie en communicatie van informatie (*grafische communicatie*). In deze studie wordt aangetoond dat grafische voorstellingen naast bovenstaande functie ook een belangrijke bijdrage kunnen leveren bij de analyse en interpretatie van informatie in beslissingsprocessen (*grafische verwerking*). Wanneer de totale hoeveelheid geregistreerde informatie zoveel mogelijk intact en zichtbaar moet blijven tijdens de analyse, zoals bij exploratief onderzoek (hypothese vormend in tegenstelling tot hypothese toetsend) of in geval de criteria en hun gewichten onduidelijk zijn (het belang van veiligheid en esthetische aspecten in vergelijking tot bijvoorbeeld bouwkosten), kunnen grafische hulpmiddelen een belangrijke ondersteuning vormen in het ontdekken en verduidelijken van onderliggende relaties in de data-structuur.

De franse cartograaf Jacques Bertin heeft een grafische methode ontwikkeld die, naar hij beweert, zowel voldoet aan de vertrouwde eisen van een correcte en overzichtelijke presentatie van gegevens, als aan de meer innoverende eisen van eenvoudige en eenduidige interactieve analyse. Deze herordenbare matrix methode (*reorderable matrix*), met name geschikt voor de analyse van multivariate gegevens, is gebaseerd op directe transformatie van meetgegevens in grafische symbolen van eenvoudige vorm, zoals stippen en vierkanten. De symbolen, waarvan de grootte correspondeert met de hoogte van de ruwe scores die ze vertegenwoordigen, worden geplaatst in de cellen van een tabel of matrix en gepresenteerd op een computer scherm. Elk van de symbolen vertegenwoordigt de score van een object (in de kolom van de matrix) op een bepaald kenmerk (in de rij). Wanneer bijvoorbeeld een huis (object) een viertal slaapkamers heeft (kenmerk), wordt deze waarde (vier) in de bijbehorende cel van de matrix geplaatst en vertaald in een daarmee corresponderend grafisch symbool.

Bij analyse van de gegevens kan de structuur van de matrix worden gewijzigd door het selecteren en vervolgens verplaatsen van rijen en kolommen. Een rechthoekmatrix leent zich uitstekend voor dergelijke, mathematisch correcte, *permutaties*. Omdat de gewijzigde matrix meteen na de verplaatsing opnieuw op het scherm wordt getekend, kan de effectiviteit van de handeling aan de hand van deze visuele feedback worden beoordeeld. Met deze visueel-interactieve methode worden

complexe cognitieve processen in de analyse van informatie geoperationaliseerd in eenvoudiger taken als visuele discriminatie, patroonherkenning en patroonvorming. In vergelijking tot computers zijn mensen erg goed in het uitvoeren van deze taken en met name voor visueel georiënteerde personen zoals bijvoorbeeld architekten en geografen lijkt dit grafisch hulpmiddel daarom zeer bruikbaar.
In deze studie worden de, aan de reorderable matrix te stellen, eisen van presentatie (correcte weergave van de originele meetgegevens waardoor een nauwkeurige interpretatie van, en verschillen, overeenkomsten en relaties tussen individuele elementen mogelijk is) en analyse (een simpele en eenduidige structurering van de informatie), experimenteel getoetst.
In Hoofdstuk 2 wordt een tweetal voorbeelden van multivariate vraagstukken, uit het toepassingsgebied van de architectuur, gegeven.
De "reorderable matrix" wordt, in Hoofdstuk 3, vergeleken met een aantal grafische methoden, ontwikkeld voor de presentatie van multivariate gegevens. De voor- en nadelen van deze methoden, bij presentatie en analyse van multivariate gegevens, worden toegelicht.
Hoofdstuk 4 is gewijd aan de verschillende grafische constructies die in de theorie van Bertin worden beschreven. Elk van deze methoden kan worden toegepast bij, en is afgestemd op, presentatie en analyse van een bepaald type van gegevens. Eén van deze, de reorderable matrix, is verder onderzocht in deze studie.
De experimentele toetsing van de verschillende criteria zijn beschreven in de hoofdstukken 5 tot en met 8.
In de Hoofdstukken 5 en 6 worden de resultaten besproken van een aantal discriminatie-, sorterings- en schattingstaken. Deze resultaten wijzen uit dat mensen deze taken vrij nauwkeurig kunnen uitvoeren. Het door Bertin aanbevolen verschil in grootte van grafische symbolen blijkt niet rechtsreeks gebaseerd op psychofysische wetmatigheden, zoals de constante uit de Weber-wet en de exponent van de machtsfunctie van Fechner (Stevens, 1957). Praktisch gezien is Bertin's aanbeveling, uit oogpunt van eenvoud en efficientie, wel bruikbaar.

Een cartografisch model, gebaseerd op Gestalt-principes, bleek in Hoofdstuk 7 de subjectief waargenomen clusters van gelijkwaardige elementen, niet optimaal te kunnen voorspellen. Visuele analyse toonde aan dat deze clusters zich op een aantal kenmerken onderscheiden van voorspelde, niet waargenomen clusters. Hierdoor is een effectiever model mogelijk door uitbreiding van het aantal variabelen.

Subjectieve beoordelingen van de mate van geordendheid van een configuratie van elementen in een matrix (*overall image*) waren zeer consistent, zowel in tijd als tussen individuen. De gestructureerdheid van een matrix lijkt hiermee een algemeen aanvaard concept. Deze subjectieve maat van geordendheid blijkt daarnaast sterk te correleren met een objectieve maat, gebaseerd op elementaire geometrische afstandsfuncties. De subjectieve impressie correspondeert hierdoor met een objectief meetbare regelmaat in de datastructuur.

In Hoofdstuk 8 wordt aangetoond dat individuen in staat zijn om, met behulp van de reorderable matrix, de structuur van een, oorspronkelijk ongeordende matrix, aanzienlijk te verbeteren. Hoewel er duidelijke verschillen tussen de individuen waren in de efficientie en snelheid waarmee ze deze opdrachten uitvoerden, benaderden allen, in hun eindoplossingen, de experimenteel bepaalde optima.

Hiermee is in deze studie aangetoond dat in het algemeen is voldaan aan de uitgangspunten of criteria van de reorderable matrix. Deze grafische methode blijkt daarmee potentieel bruikbaar in de analyse van gegevens in multivariate beslissingsprocessen. Gezien het experimenteel karakter van een aantal onderzoeken dat binnen deze studie is verricht, is nader onderzoek naar de draagwijdte van dit instrument in uiteenlopende praktisch situaties, gewenst.

About the author

Wim Adams was born on May 22, 1958 in Tilburg. After completing his VWO-exam at the Paulus Lyceum, he studied Psychology from 1977 until 1985 at the Tilburg University. From 1985 until 1986 he was employed as a research assistant in the department of Ergonomic and Educational Psychology at the Tilburg University. Since 1986 he is a research assistant in the department of Architecture, Building and Planning of the Eindhoven University of Technology.

Stellingen behorend bij het proefschrift:

Supporting Decision-Making Processes:
A Graphical and Interactive Analysis of Multivariate Data

W.T.C.F. Adams

1. De ontwikkeling van grafische technieken voor de presentatie en analyse van gegevens is sterk achtergebleven bij de ontwikkeling van mathematisch-statistische technieken. Dit is in belangrijke mate veroorzaakt door het gebruik van computers bij de laatstgenoemde manier van bewerking van gegevens en niet door het verschil in kwaliteit tussen de technieken (Fienberg, 1979; Schmid, 1983).

Fienberg, S.E.
Graphical Methods in Statistics. The American Statistician, 1979, 33, pp. 165-178.

Schmid, C.F.
Statistical Graphics. John Wiley, NY, 1983.

2. Het door Wainer & Thissen (1981) gehanteerde begrip "analysis" in hun "Graphical Data Analysis" is onzorgvuldig. Hoewel sommige van de door hen besproken grafisch-analytische methoden inderdaad de mogelijkheid bieden tot een meer diepgaande exploratie van de gepresenteerde gegevens en de structuur van hun onderlinge relaties, zijn andere beperkt tot de communicatie of illustratie van enkele, specifieke resultaten of kenmerken.

Wainer, H. & Thissen, D.
Graphical Data Analysis. Annual Review of Psychology, 1981, vol. 32, 191-241.

3. Het toegankelijker worden van geavanceerde statistische analyse technieken werkt het onzorgvuldig gebruik ervan in de hand.

4. Wanneer men grafische methoden van weergave wil gebruiken om gegevens te analyseren, zijn methoden die weinig voorbewerking vereisen, zoals de "herordenbare matrix methode" (Bertin, 1981), meer geschikt dan methoden die veel voorbewerking vereisen, zoals de "Kleiner-Hartigan tree" symbolen (Chambers, Cleveland, Kleiner & Tukey, 1983).

Bertin, J.
Graphics and Graphic Information Processing W. de Gruyter, Berlin. New York, 1981.

Chambers, J.M., Cleveland, W.S., Kleiner, B. & Tukey, P.A.
Graphical Methods for Data Analysis. Wadsworth, Belmont, CA., 1983.

5. In het interaktief herordenen van een grafische data-matrix is het bij het positioneren van de grafische elementen mogelijk om zogenaamde

extrinsieke informatie te gebruiken (informatie over de geregistreerde objecten die niet is opgenomen in de data-matrix maar vaak wel bij de gebruiker bekend is, Bertin, 1981). Hoewel deze mogelijkheid niet altijd zal leiden tot duidelijker herkenbare patronen van grafische elementen geeft hij wel informatie over specifieke beslissingen van de gebruiker in het ordeningsproces. De herkenbaarheid hiervan maakt het beslissingsproces bespreekbaar.

6. Een computerprogramma dient vooral beoordeeld te worden op zijn gebruikersvriendelijkheid, het gemak waarmee de gebruiker met het programma zijn doelen kan realiseren, en minder op zijn gebruiksvriendelijkheid, het gemak waarmee de gebruiker de mogelijkheden van het programma kan benutten.

7. Op door gebruikers te volgen routes kunnen naast keuzepunten (punten op een route waarop een keuze over de voortzetting van de route mogelijk is) ook beslispunten (punten op een route waarop de gebruiker een bewuste afweging maakt over de voortzetting van de route) worden onderscheiden. Indien de keuzepunten en de beslispunten op een route niet samenvallen, verdient het de voorkeur het aanbieden van verwijzingsinformatie te concentreren op de beslispunten en minder op de keuzepunten (Joanknecht en Venemans, 1985).

Joanknecht, J.W. & Venemans, P.J.
Hoezo Bewegwijzering? Een inventarisatie van de problemen met de bewegwijzeringssituatie van een groot ziekenhuis. Intern rapport, Katholieke Universiteit Brabant, 1985.

8. Zolang een groot aantal mensen moeite heeft met het lezen en begrijpen van plattegronden, is de functie van langs de straat opgestelde stads- en wijkplattegronden er vooral een van expressionistische stadsverfraaiing.