INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.
Tullis, Thomas Stuart

PREDICTING THE USABILITY OF ALPHANUMERIC DISPLAYS

Rice University

University Microfilms International

Copyright 1984 by Tullis, Thomas Stuart

All Rights Reserved
PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark √.

1. Glossy photographs or pages ______
2. Colored illustrations, paper or print ______
3. Photographs with dark background ______
4. Illustrations are poor copy ______
5. Pages with black marks, not original copy ______
6. Print shows through as there is text on both sides of page ______
7. Indistinct, broken or small print on several pages √
8. Print exceeds margin requirements ______
9. Tightly bound copy with print lost in spine ______
10. Computer printout pages with indistinct print ______
11. Page(s) _________ lacking when material received, and not available from school or author.
12. Page(s) _________ seem to be missing in numbering only as text follows.
13. Two pages numbered _________. Text follows.
14. Curling and wrinkled pages ______
15. Other _______________
RICE UNIVERSITY

PREDICTING THE USABILITY OF ALPHANUMERIC DISPLAYS

by

THOMAS STUART TULLIS

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE DOCTOR OF PHILOSOPHY

APPROVED, THESIS COMMITTEE:

David M. Lane, Associate Professor of Psychology
Chairman

William C. Howell, Professor of Psychology

Richard J. Stoll, Assistant Professor of Political Science

HOUSTON, TEXAS

MAY, 1984
Copyright
Thomas Stuart Tullis
1984
PREDICTING THE USABILITY OF ALPHANUMERIC DISPLAYS

by

THOMAS STUART TULLIS

ABSTRACT

A review of the literature on alphanumeric displays, especially computer-generated displays, suggests that four basic characteristics of display formats affect how well users can extract information from the displays: (1) Overall density—the number of characters displayed, expressed as a percentage of the total spaces available; (2) Local density—the number of other characters near each character; (3) Grouping—the extent to which characters on the display form well-defined perceptual groups; and (4) Layout complexity—the extent to which the arrangement of items on the display follows a predictable visual scheme. Objective ways of measuring these display characteristics have been developed and implemented in a computer program.

In Experiment 1, 520 computer-generated displays that varied on these display measures were studied. Search times to locate data items on the displays were measured as well as subjective ratings of ease of use. Regression equations were developed to predict the search times and subjective ratings using the display measures. The results indicated that both search times \((R^2 = .508)\) and subjective ratings \((R^2 = .805)\) could be predicted quite well.

In Experiment 2, the regression equations developed in Experiment 1 were used to predict, \textit{a priori}, search times and subjective
ratings for a new set of 150 displays. The regression equations generalized quite well, resulting in high correlations between predicted and actual search times \((r = .800)\) and subjective ratings \((r = .799)\).

The regression equations indicate that the most important predictors of search time are two measures associated with the grouping of characters: the number of groups on the display and the average visual angle subtended by those groups. Likewise, the most important predictors of subjective rating are a measure of local density, which is essentially how "tightly packed" the display is, and a measure of layout complexity, which is essentially how well the items on the display are aligned with each other.
# TABLE OF CONTENTS

1. INTRODUCTION ................................................. 1
   1.1 Scope of the Research .................. 1
   1.2 Impact of Display Design ............. 2
   1.3 Goals of the Research ................ 5
2. METHOD OF THE LITERATURE REVIEW .......... 7
   2.1 Guidelines ........................................ 7
   2.2 Empirical Studies ......................... 9
   2.3 Selection of Display Characteristics .. 15
3. CHARACTERISTICS OF DISPLAYS ............ 17
   3.1 Overview ....................................... 17
   3.2 Overall Density ............................... 19
   3.3 Local Density ................................. 24
   3.4 Grouping ...................................... 35
   3.5 Layout Complexity ......................... 46
   3.6 Summary ...................................... 58
4. EXPERIMENT 1: DEVELOPMENT OF THE PREDICTION SYSTEM . 62
   4.1 Introduction .................................. 62
   4.2 Method ....................................... 66
   4.3 Results ....................................... 74
   4.4 Discussion ................................... 88
5. EXPERIMENT 2: VALIDATION OF THE PREDICTION SYSTEM ...... 92
   5.1 Introduction .................................. 92
   5.2 Method ....................................... 92
   5.3 Results ....................................... 99
   5.4 Discussion ................................... 105
6. GENERAL DISCUSSION: APPLYING THE PREDICTION SYSTEM TO DISPLAYS IN THE LITERATURE .................................................. 109
7. CONCLUSIONS ........................................................................... 137
   REFERENCES .............................................................................. 145
   APPENDIX A ........................................................................... 152
LIST OF FIGURES

Figure 1. Display of narrative format from Tullis (1981) .................................................. 4
Figure 2. Display of structured format from Tullis (1981) .................................................. 4
Figure 3. Character spaces on CRT screen subtended by 5-deg visual angle and their approximate weights ............ 31
Figure 4. An empty display .................................................. 33
Figure 5. A completely filled display .................................................. 33
Figure 6. A display with the left half filled .................................................. 34
Figure 7. A display with half the spaces filled uniformly across the display .............................. 34
Figure 8. Sample display to illustrate grouping algorithm .................................................. 42
Figure 9. Sample display with lines connecting all pairs of characters separated by less than a threshold distance .................................................. 42
Figure 10. Narrative format from Tullis (1981) with grouping indicated .................................. 44
Figure 11. Structured format from Tullis (1981) with grouping indicated ................................ 44
Figure 12. Original catalog page from Bonsiepe (1968) .................................................. 49
Figure 13. Redesigned catalog page from Bonsiepe (1968) .................................................. 49
Figure 14. Original catalog page from Bonsiepe (1968) showing rectangles defining each group .................................................. 50
Figure 15. Redesigned catalog page from Bonsiepe (1968) showing rectangles defining each group .................................................. 50
Figure 16. Original version of display from Stewart (1976) with rectangles added to indicate labels and data items .................................................. 55
Figure 17. Redesigned version of display from Stewart (1976) with rectangles added to indicate labels and data items .................................................. 55
Figure 18. Narrative format from Tullis (1981) with labels and data items inscribed by rectangles................. 57
Figure 19. Structured format from Tullis (1981) with labels and data items inscribed by rectangles............... 57
Figure 20. Sample output from program........................ 64
Figure 21. Two examples of airline displays used in Experiment 1................................................. 67
Figure 22. Two examples of motel displays used in Experiment 1.................................................... 68
Figure 23. Experiment 1: Scatterplot of search time and size of groups.............................................. 77
Figure 24. Experiment 1: Plot of $R^2$ for multiple regressions using different numbers of display measures to predict search time................................................. 78
Figure 25. Experiment 1: Plot of actual search times versus search times predicted from a multiple regression using overall density, local density, number of groups, and size of groups.......................... 82
Figure 26. Experiment 1: Scatterplot of subjective rating and item uncertainty...................................... 84
Figure 27. Experiment 1: Plot of $R^2$ for multiple regressions using different numbers of display measures to predict subjective rating................................................. 85
Figure 28. Experiment 1: Plot of actual subjective ratings versus subjective ratings predicted from a multiple regression using overall density, local density, number of groups, size of groups, number of items, and item uncertainty.......................... 89
Figure 29. Two examples of book displays used in Experiment 2............................................................ 93
Figure 30. Experiment 2: Scatterplot of actual search times and search times predicted from a model developed in Experiment 1 that uses overall density, local density, number of groups, and size of groups. Letters indicate different displays.................. 100
Figure 31. Experiment 2: Scatterplot of actual subjective ratings and subjective ratings predicted from a regression equation developed in Experiment 1 that uses overall density, local density, number of groups, size of groups, number of items, and item uncertainty. Letters indicate different displays. ........................................ 103

Figure 32. Experiment 2: Scatterplot of actual transformed scores and predicted transformed scores based on regression equations developed in Experiment 1. Transformed scores are the sum of z-score transformations for search times and subjective ratings. Letters indicate different displays........................................ 108

Figure 33. 30% density display studied by Dodson and Shields (1978)........................................ 114

Figure 34. 50% density display studied by Dodson and Shields (1978)........................................ 114

Figure 35. 70% density display studied by Dodson and Shields (1978)........................................ 115

Figure 36. Plot of actual and predicted search times from Dodson and Shields (1978) as a function of display density........................................ 118

Figure 37. Example of display studied by Ringel and Hammer (1964)........................................ 120

Figure 38. Actual and predicted search times for main effect of Number of Lines in Ringel and Hammer (1964) study........................................ 122

Figure 39. Actual and predicted search times for main effect of Density in Ringel and Hammer (1964) study........................................ 123

Figure 40. Example of a 19-display-item format studied by Callan, Curran, and Lane (1977)........................................ 124

Figure 41. Actual and predicted search times from Callan, Curran, and Lane (1977)........................................ 126

Figure 42. First pair of 'before' (top) and 'after' (bottom) displays presented by Marcus (1982)........................................ 129

Figure 43. Second pair of 'before' (top) and 'after' (bottom) displays presented by Marcus (1982)........................................ 130

Figure 44. Third pair of 'before' (top) and 'after' (bottom) displays presented by Marcus (1982)........................................ 131
Figure 45.  'Bad' version of display for a hypothetical hotel reservation and billing system from Pakin and Wray (1982).............................. 133

Figure 46.  First of two new displays presented by Pakin and Wray (1982) to replace the display in Figure 45................................. 134

Figure 47.  Second of two new displays presented by Pakin and Wray (1982) to replace the display in Figure 45................................. 134

Figure 48.  Display with one large group of characters........ 139

Figure 49.  Display with 240 small groups of characters.... 139

Figure 50.  Display with 32 intermediate-size groups of characters.......................... 140

Figure 51.  Tightly packed version of Figure 50........ 143

Figure 52.  Poorly aligned version of Figure 50........... 143
LIST OF TABLES

TABLE 1. Guidelines Addressing Alphanumeric Display Design .................................................. 8

TABLE 2. Categorization of Empirical Studies on Displays by Type of Display and Type of Task .......... 10


TABLE 4. Means, Standard Deviations, and Ranges of Six Display Measures for 520 Displays Used in Experiment 1 ................................................................. 69

TABLE 5. Correlations among Six Display Measures for 520 Displays Used in Experiment 1 ............ 71

TABLE 6. Examples of Questions Asked for Airline and Motel Displays in Experiment 1 ................ 73

TABLE 7. Experiment 1: Simple Correlations of Six Display Measures with Search Times and Subjective Ratings ................................................................. 76

TABLE 8. Experiment 1: Best-fitting Multiple Regressions Predicting Search Time at Each Number of Predictor Variables from Two to Six ................................. 79

TABLE 9. Experiment 1: Best-fitting Multiple Regressions Predicting Subjective Rating at Each Number of Predictor Variables from Two to Six .......................... 86

TABLE 10. Means, Standard Deviations, and Ranges of Six Display Measures for 150 Displays Used in Experiment 2 ................................................................. 95

TABLE 11. Examples of Questions Asked for Book Displays in Experiment 2 ................................ 98

TABLE 12. Simple Correlations Between Actual Search Times from Experiment 2 and Predicted Search Times Based on Five Different Regression Equations from Experiment 1 .................................................. 102

TABLE 13. Simple Correlations Between Actual Subjective Ratings from Experiment 2 and Predicted Subjective Ratings Based on Five Different Regression Equations from Experiment 1 .................................................. 106
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>Results of Applying the Prediction System to Search Time Data in the Literature</td>
<td>110</td>
</tr>
<tr>
<td>15</td>
<td>Values of Six Display Parameters for Displays Studied by Dodson and Shields (1978)</td>
<td>116</td>
</tr>
<tr>
<td>16</td>
<td>Predicted Search Times and Subjective Ratings for 'Bad' and 'Good' Versions of Displays Published in the Literature</td>
<td>128</td>
</tr>
<tr>
<td>17</td>
<td>Values of Six Display Variables, Predicted Search Time, and Predicted Subjective Rating for Five Prototypical Displays</td>
<td>141</td>
</tr>
</tbody>
</table>
PREDICATING THE USABILITY OF ALPHANUMERIC DISPLAYS

1. INTRODUCTION

People extract information from alphanumeric displays every day. They scan newspapers, read highway signs, look up telephone numbers, or read the latest bestseller. In addition, a growing number of people spend most of their days working with computer-generated alphanumeric displays. In each of these situations the individual is presented with a frame of information, such as a newspaper page or a CRT screen, that must be processed according to the demands of the task. Each of these frames is somehow formatted to aid the viewer in processing the information. One of the purposes of this research is to determine what characteristics of these formats have an effect on the user's ability to process the information.

1.1 Scope of the Research

To reduce this topic to a manageable size, some restrictions on its scope must be adopted. This research will focus on the following types of displays:

1. Computer-generated displays, especially CRT displays. The primary reason for focusing on computer-generated displays is their widespread use and the need to understand their effect on human performance. Many of the topics addressed, however, will be equally relevant to other types of displays.
2. **Alphanumeric displays.** This research is restricted to alphanumeric displays primarily because most computer-generated displays in commercial use are alphanumeric. This is not intended to deny the importance of graphic displays, which have been shown to be beneficial to human performance under certain circumstances (e.g., Tullis, 1981).

3. **Monochromatic displays.** As with alphanumeric displays, the main reason for focusing on monochromatic displays is simply that most of the computer-generated displays currently in use are monochromatic. In addition, the use of color in displays has been discussed extensively elsewhere (e.g., Carter, 1982; Carter & Cahill, 1979; Christ, 1975; Christ & Corso, 1983).

4. **Formatted displays.** This research will focus on formatted displays (Rosenthal, 1979) as opposed to narrative displays. The primary difference is that the viewer's task with a formatted display is usually not reading the frame from left-to-right and top-to-bottom. Instead, the task generally involves searching and selectively encoding parts of the display (e.g., searching a computer-displayed menu for the desired function).

1.2 Impact of Display Design

Placing these restrictions on the type of displays to be studied may give the impression that the scope has been reduced to an almost trivial problem. However, the number of such displays in use and the amount of time spent in working with them is overwhelm-
ing. Galitz (1980, pp. 105-106) reported the results of a study which projected that a single computer system for an insurance company would process 4.8 million screens per year. When analogous estimates are made for computer systems in an organization the size of the Bell System, the numbers become staggering. For example, in only one system, the Automated Repair Service Bureau (ARSB), it can be conservatively estimated that Bell System employees using the system must extract information from 344 million distinct frames (screens, printouts) per year.¹ The importance of designing these frames effectively becomes clear when one considers that if users took only 1 second longer to extract information from each frame, that would require an additional 55 person-years just to extract the information.²

Can the design of such a display actually make a difference in how rapidly the user extracts information from it? Consider the displays shown in Figures 1 and 2, taken from a study evaluating alternative formats for one of the CRT displays in the ARSB system (Tullis, 1981). These are displays of results from computerized tests of a telephone line. Figure 1 shows a "narrative" format originally designed to convey the test results. Figure 2 shows a

---

¹ This estimate is based on the number of telephone lines served by the ARSB as of December, 1981, the average trouble report rate per line, and the average number of times that employees must extract information from screens or printouts for each report.

² Assuming 6.3 million sec per person-year, which is based on the number of working hours per person per day, and the number of working days per year.
TEST RESULTS SUMMARY: GROUND

GROUND, FAULT T-G
3 TERMINAL DC RESISTANCE
  > 3500.00 K OHMS T-R
  = 14.21 K OHMS T-G
  > 3500.00 K OHMS R-G
3 TERMINAL DC VOLTAGE
  = 0.00 VOLTS T-G
  = 0.00 VOLTS R-G
VALID AC SIGNATURE
3 TERMINAL AC RESISTANCE
  = 8.82 K OHMS T-R
  = 14.17 K OHMS T-G
  = 628.52 K OHMS R-G
LONGITUDINAL BALANCE POOR
  = 39 DB
COULD NOT COUNT RINGERS DUE TO LOW RESISTANCE
VALID LINE CKT CONFIGURATION
CAN DRAW AND BREAK DIAL TONE

Figure 1. Display of narrative format from Tullis (1981)

*******************************************************************************
*                        * 
*            TIP GROUND        14 K            * 
*                        * 
*******************************************************************************

DC RESISTANCE DC VOLTAGE AC SIGNATURE
3500 K T-R 0 V T-G 9 K T-R
14 K T-G 0 V T-G 14 K T-G
3500 K R-G 0 V R-G 629 K R-G

BALANCE CENTRAL OFFICE
39 DB VALID LINE CKT
DIAL TONE OK

Figure 2. Display of structured format from Tullis (1981)
"structured" format in which the test results were redesigned using a variety of techniques (eliminating unnecessary information, grouping related data, etc.). An evaluation was conducted in which Bell System employees were trained on the interpretation of these displays and then tested using questions of the type they would have to answer for themselves on the job. The evaluation showed that, after practice, mean time to answer a question about the display was 8.3 sec for the "narrative" format and 5.0 sec for the "structured" format. That savings of 3.3 sec per display translates to 79 person-years saved.

1.3 Goals of the Research

Clearly, the design of a display can have a significant impact on the user's ability to extract information from it. The question that remains is why are there such differences in the usability of displays? For example, what characteristics of the display shown in Figure 2 make it easier to use than the display shown in Figure 1? The primary goals of this research, then, are as follows:

1. To identify, through a review of the literature, those characteristics of alphanumeric displays that appear to affect the user's ability to extract information from the displays

3. Derived from: the number of telephone lines served by the computerized testing system as of December, 1981; the average trouble report rate per line per year; the average number of test results displays interpreted per trouble report; a 3.3 sec savings per display; and 6.3 million sec per person-year.
2. To develop a computer-based system for objectively measuring those characteristics

3. To experimentally determine regression equations that use those characteristics to predict search times and subjective ratings of displays

4. To validate those regression equations experimentally and through application to data published in the literature
2. METHOD OF THE LITERATURE REVIEW

The inter-disciplinary nature of this topic required searching a wide variety of literature sources for relevant information. This included, but was not limited to, the human factors and basic psychological literature, as well as the computer literature. The survey covered the literature from 1960 through 1982. The literature on displays may be divided into two general categories: (1) Guidelines, and (2) Empirical studies.

Several existing reviews of the literature on human factors in computer systems were particularly useful in this survey. Ramsey and Atwood (1979) published an extensive review of human factors in computer systems along with an associated bibliography of that literature compiled by Ramsey, Atwood, and Kirshbaum (1978). Ramsey and Atwood (1979) provide a particularly good overview of the research on "Informational Properties of Displays" (pp. 119-129) that provided a starting point for this review. More recently, Williges and Williges (1981) published a compilation of user considerations in computer-based information systems.

2.1 Guidelines

Table 1 shows that the guidelines commonly take on one of two forms: either highly specific lists of rules for information display (e.g., Engel & Granda, 1975; Smith, 1980, 1981, 1982), or more conceptual discussions of display design in general (e.g., Cakir, Hart, & Stewart, 1980; Peterson, 1979). One often encounters problems in applying these guidelines. The lists of rules are commonly too specific to a particular type of display or
TABLE 1. Guidelines Addressing Alphanumeric Display Design

<table>
<thead>
<tr>
<th>Formal lists of rules for display design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bailey (1982)</td>
</tr>
<tr>
<td>Engel and Granda (1975)</td>
</tr>
<tr>
<td>Galitz (1980)</td>
</tr>
<tr>
<td>NASA (1980)</td>
</tr>
<tr>
<td>Pew and Rollins (1975)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Informal discussions of display design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bonsiepe (1968)</td>
</tr>
<tr>
<td>Cakir, Hart, and Stewart (1980)</td>
</tr>
<tr>
<td>Cropper and Evans (1968)</td>
</tr>
<tr>
<td>Danchak (1976)</td>
</tr>
<tr>
<td>Grace (1966)</td>
</tr>
<tr>
<td>Green (1976)</td>
</tr>
<tr>
<td>Jones (1978)</td>
</tr>
<tr>
<td>Marcus (1981, 1982)</td>
</tr>
<tr>
<td>Mehlmann (1981)</td>
</tr>
<tr>
<td>Miller and Thomas (1977)</td>
</tr>
<tr>
<td>Pakin and Wray (1982)</td>
</tr>
<tr>
<td>Peterson (1979)</td>
</tr>
<tr>
<td>Rosenthal (1979)</td>
</tr>
<tr>
<td>Stewart (1976)</td>
</tr>
<tr>
<td>Uber, Williams, Hisey, and Siekert (1968)</td>
</tr>
</tbody>
</table>
system to be applicable to one's own problem; or they address the
details of a display (e.g., "Use a 'MM/DD/YY' format for dates")
without adequately addressing the display as a whole. On the other
hand, the more conceptual guidelines are often too general to be
directly applicable in the design of a given display (e.g., "The
format should be simple and straight-forward"). What is lacking is
a "middle level" of guidelines that are general enough to be appli-
cable to a variety of displays but specific enough to be directly
usable. Part of the goal of this review is to search out those
existing guidelines that fall within this middle level.

Another major problem with the guidelines is that they are not
necessarily based on empirical data. More often than not they are
based on the authors' own ideas about display design, and only
occasionally supported by references to empirical data. However,
one would hope that these ideas are based on extensive experience
in designing and using displays. As such, there may be useful
information in most of these guidelines, including those that have
no supporting empirical data, since they point to possibly fruitful
areas of research.

2.2 Empirical Studies

Table 2 categorizes the empirical studies related to display
design in two ways: by the type of display and the type of task.
In general, two types of displays have been studied: (1) simple,
artificial displays, and (2) more complex, realistic displays.
Even though this review focuses on the particular type of displays
outlined earlier (computer-generated, alphanumeric, monochromatic,
<table>
<thead>
<tr>
<th>TABLE 2. Categorization of Empirical Studies on Displays by Type of Display and Type of Task</th>
</tr>
</thead>
</table>

**Simple, Artificial Displays** (e.g., single character targets)

**Visual Search**
- Atkinson, Holmgren, and Juola (1969)
- Banks and Prinzmetal (1976)
- Brown and Monk (1975)
- Burns (1979)
- Egeth, Atkinson, Gilmore, and Marcus (1973)
- Mackworth (1976)
- Treisman (1982)

**Complex, Realistic Displays** (e.g., words, data, system displays)

**Visual Search**
- Callan, Curran, and Lane (1977)
- Card (1982)
- Dodson and Shields (1978)
- Vartabedian (1971)
- Williams (1966)
- Woodward (1972)

**Question-Answering**
- Coffey (1961)
- Grace (1966)
- Ringel and Hammer (1964)
- Tullis (1981)
- Wright (1968, 1977)

**Problem-Solving/Decision-Making**
- Baker and Goldstein (1966)
Cicchinelli and Lantz (1978)
Dorris, Sadosky, and Connolly (1977)
Landis, Slivka, and Jones (1967)
Silver, Jones, and Landis (1966)

Reading
Kolers, Duchnicky, and Ferguson (1981)
Poulton and Brown (1968)
Tinker (1955)

Subjective Ratings
Christie (1981)
Siegel and Fischl (1971)
Vitz (1966)
formatted), some empirical studies using displays of other types (e.g., graphic, color, text) have been included when their results are relevant to the topic of interest.

2.2.1 Simple, artificial displays

The simpler displays usually involve arrays of individual letters or numbers (e.g., Atkinson, Holmgren, & Juola, 1969) or individual symbols such as dots and squares (e.g., Mackworth, 1976). The number of items in each display is typically small (e.g., 1-5 for the Atkinson et al. study), although some do involve larger numbers (e.g., up to 81 for Treisman’s, 1982, Experiment 2). Most studies using the simpler displays involve a visual search task. Typically, subjects are instructed to indicate the presence or absence of a target in each display (e.g., Egeth, Atkinson, Gilmore, & Marcus, 1973) or to indicate what target from a predefined set appeared (e.g., Banks & Prinzmetal, 1976).

The studies listed in Table 2 under simple displays represent only a sample of the studies that have been conducted in the visual search tradition. (For reviews of visual search see Teichner & Krebs, 1974, and Teichner & Mocarnuk, 1979.) By virtue of the simplicity of the displays, these studies usually address well defined issues (e.g., effect of number of elements) in a closely controlled way. As such, their relevance to the real-world use of displays is often questionable, but they at least suggest effects that may exist for real-world displays.
2.2.2 Complex, realistic displays

The more complex displays usually involve arrays of words, numeric data, or alphanumeric codes. Often, these studies use displays proposed for a computer system (e.g., Dodson & Shields, 1978; Grace, 1966; Tullis, 1981). The number of items in each display is typically large (e.g., 100 to 250 items in the Ringel & Hammer, 1964, study) and there are often complex relationships among the items (e.g., currency conversion tables and travel timetables used by Wright, 1968 and 1977).

The greater complexity of these displays allows for a greater diversity of tasks to be performed using them. Table 2 divides the studies using these displays into five categories according to the type of task:

1. Visual Search. As with the simple displays, a common task is to report a well-defined item on the display. In some cases, the subject is told the identity of an item and then simply indicates its location (e.g., Card, 1982). In other cases, the subject is told some label on the display and then reports the data associated with that label (e.g., Callan, Curran, & Lane, 1977).

2. Question-Answering. Another common technique is to ask subjects a variety of questions about the displays. These range from simple questions that can be answered by retrieving only one item from the display (thus reducing to a simple search task) to complex questions involving the integration of numerous items on the display. For example, Ringel and
Hammer (1964), using displays of the status of military units, asked questions like, "Which unit has an armor equipment status of 95 and is combat experienced?". This type of question required subjects to find a specific conjunction of characteristics in two separate columns of a table.

3. **Problem-Solving/Decision-Making.** Another technique is to pose a problem to the subjects and have them solve that problem using the displays. Sometimes the problem can be solved using only one display (e.g., Cicchinelli & Lantz, 1978) and sometimes it requires a sequence of displays (e.g., Baker & Goldstein, 1966). The important distinction between this technique and the previous one (question-answering) is that problem-solving always involves the same problem on all displays. For example, Silver, Jones, and Landis (1966) described a hypothetical trucking company to their subjects, who were then given a complex display describing the current status of the company's truck loads, drivers, and trucks. Given the destination for each of the loads, the subjects were instructed to assign loads to drivers and trucks to keep costs at a minimum.

4. **Reading.** A few studies using a reading task manipulated variables of interest to the design of formatted displays. Typically, subjects in such studies are instructed to read the text and then are tested for comprehension. Some studies (e.g., Kolers, Duchnicky, & Ferguson, 1981) monitor eye movements to determine the pattern of fixations.
5. **Subjective Ratings.** In some studies, the subjects' primary task is to make qualitative ratings of the displays. Subjects have been instructed to rate each display on a scale of overall quality (Christie, 1981, pp. 202-205), to rate the similarity of pairs of displays (Siegel & Fischl, 1971), or to rank order their preferences for a set of displays (Vitz, 1966).

2.3 Selection of Display Characteristics

The literature was studied to determine what underlying characteristics of displays are addressed by either the guidelines or the empirical studies. The following criteria for defining and selecting these underlying characteristics were adopted:

1. The characteristic must be related to the *spatial array* of characters on the display. This eliminates a variety of highlighting techniques (e.g., blinking, increased intensity, reverse video) that clearly affect the user's processing of the display. These were eliminated simply because they represent a qualitatively different aspect of display design. Blink coding in particular has been discussed in detail by Smith and Goodwin (1971, 1972). Characteristics related to the legibility of individual characters have also been eliminated. The effects on character legibility of such factors as luminance contrast, dot matrix size, and raster scan resolution have been discussed thoroughly by Shurtleff (1980).

2. The characteristic must be *objectively defined*, or at least
have the potential for it. The main criterion used for deciding this was to ask, "Could a computer program be written to assess this characteristic?". If the answer was either "Yes" or "Probably," the characteristic was included. This criterion eliminates a wide variety of guidelines that are dependent on semantic or contextual information (e.g., presenting information in a "logical" sequence according to its semantic content). The elimination of these semantic characteristics is not meant to deny their importance: they clearly are critical to the usability of a display. However, semantic characteristics represent a different realm from the spatial characteristics considered in this paper.

3. The characteristic must be applicable to any alphanumeric display. This eliminates many guidelines that apply only to certain types of displays (e.g., guidelines addressing the format of specific items, such as a date, that not all displays contain).
3. CHARACTERISTICS OF DISPLAYS

3.1 Overview

By applying the preceding criteria to the guidelines and empirical data, four basic characteristics of alphanumeric display formats were identified. These four characteristics, some of which are better-defined than others, are as follows:

1. **Overall density** - the number of characters displayed, often expressed as a percentage of the total character spaces available.

2. **Local density** - the number of filled character spaces near each character, often manipulated by altering line spacing.

3. **Grouping** - the extent to which items form well-defined perceptual groups.

4. **Layout Complexity** - the extent to which the arrangement of items on the frame follows a predictable visual scheme.

These four characteristics are not necessarily independent of each other. For example, in the set of real-world displays, overall density and local density have a high positive correlation. Likewise, increasing the overall density tends to decrease the extent to which items can be arranged into distinctly separate groups.

This derivation of characteristics from the guidelines and empirical data was not always a well-defined process. Specifically, there were some situations where a new characteristic could
have been defined but was not for a variety of reasons. For example, several studies (Coffey, 1961; Williams, 1966; Woodward, 1972) have investigated the effect of presenting data items in a horizontal vs. vertical orientation. The consensus seems to be that items in a horizontal orientation can be encoded more readily than those in a vertical orientation. Such a characteristic of "orientation of items" was not included primarily because it seems to have little practical significance, since almost all displays are designed with elements arranged horizontally (due to reading biases). The lack of practical significance of this finding is also reflected by the fact that only one of the discussions of guidelines mentions that a horizontal arrangement is preferred (Danchak, 1976, p. 34).

Another example of a characteristic that could have been included is the letter case (upper or lower) of items on the display. Numerous authors have proposed that normal upper- and lower-case should be used for presentation of text while upper-case alone should be used for labels and visual search tasks (Engel & Granda, 1975, p. 17; Galitz, 1980, p. 111; Mehlmann, 1981, p. 118; Smith, 1982, p. 97). These guidelines are based on studies by Poulton and Brown, 1968, and Tinker, 1955, which showed that combined upper- and lower-case text was read about 13% faster than text in all capitals, and a study by Vartabedian (1971) which showed that search time to find a word on a CRT display was about 13% shorter for upper-case words than for lower-case words. The characteristic of "letter case" was not included because it is only indirectly related to the spatial array of characters on the screen. One could argue that letter case has an effect on the spa-
tial array since upper-case letters are, on the average, larger than lower-case letters, but this seems to be stretching the point.

One last example of a characteristic that could have been included is the consistency of related displays. Numerous guidelines have advocated the formatting of items on related displays in a consistent manner: "When using the same items on different displays, make an attempt to locate them in the same place on all displays" (Bailey, 1982, p. 133); "The ordering and layout of corresponding data fields should be consistent from one display to the next" (Smith, 1982, p. 104). Although such consistency is clearly important, it was not included in this analysis because it represents a level of organization beyond a single display. Consistency is only defined on the basis of a set of displays.

Each of the four characteristics of displays will now be considered in detail. The discussion of each characteristic will first cover the relevant guidelines, followed by the empirical data, and then a synthesis of the two sets of information. Much of the discussion will center on ways of measuring these characteristics.

3.2 Overall Density

3.2.1 Guidelines

A common theme in many of the display design guidelines is that only "relevant" information should be displayed (e.g., Cakir et al., 1980, p. 114; Galitz, 1980, p. 108), or that the display should not appear "cluttered" (e.g., Green, 1976, p. 145; Peterson, 1979, p. 20). These guidelines are addressing the same basic
concept: the total amount of information displayed on a single frame should be kept to a minimum. Some of the authors go on to rationalize this guideline by claiming that high information density causes "the human perceptual channels [to] become overloaded" (Cropper & Evans, 1968, p. 96), "leads to confusion and an increased error rate" (Engel & Granda, 1975, p. 8), causes "greater...competition among screen components for a person's attention" (Galitz, 1980, p. 108), causes a "psychological strain" (Green, 1976, p. 145), or "increases search time" (Stewart, 1976, p. 142).

Three of the authors take this idea a step further and discuss specific values for overall density. Danchak (1976) states, "Experience shows that display-loading (the percentage of active screen area) should not exceed 25 percent" (p. 33). He goes on to state that "an analysis of existing CRT displays that were qualitatively judged 'good' revealed a loading on the order of 15 percent" (p. 33).

In a set of guidelines for the design of Spacelab displays, NASA (1980) states that "density generally should not exceed 60% of the available character spaces" (p. 3-26). They support this guideline by claiming that "empirical tests...indicate that response time and accuracy begin to degrade rapidly when data presented is above 60% density" (p. 3-26). Apparently this refers to the Dodson and Shields (1978) study of Spacelab displays. They manipulated density using values of 30%, 50%, and 70% and found increases in search time with increasing density. However, the support for the assertion that performance begins to degrade rapidly above 60% den-
sity is not clear, since Dodson and Shield’s (1978) data appear to be fit reasonably well by a simple linear function over the range they studied.

Smith (1980, 1981, 1982) does not provide a specific upper limit for display density, but in a checklist for "data display" he refers to three levels of density for tabular data and data forms (1982, p. 53):

- high (> 600 char)
- moderate
- low (< 300 char)

Assuming a standard 24 x 80 character CRT, these values correspond to the following densities:

- high (> 31.2 %)
- moderate
- low (< 15.6 %)

Interestingly, Smith’s threshold for high density of 31.2% is not too far from Danchak’s density limit of 25%. Likewise, Smith’s threshold for low density of 15.6% is essentially the same as Danchak’s optimal density of 15%.

3.2.2 Empirical Data

Studies have repeatedly shown that human performance deteriorates with increasing display density. The most common finding of visual search studies is that increasing the number of displayed items increases time and errors in locating the target.
This has been shown in search studies using simple displays (e.g., Atkinson et al., 1969; Burns, 1979; Egeth et al., 1973; Mackworth, 1976; Treisman, 1982) as well as complex displays (Callan et al., 1977; Dodson & Shields, 1978). Similar results have been found in question-answering studies (Coffey, 1961; Ringel & Hammer, 1964) and problem-solving studies (Baker & Goldstein, 1966; Cicchinelli & Lantz, 1978; Dorris, Sadosky, & Connolly, 1977; Landis, Slivka, & Jones, 1967).

3.2.3 Synthesis

The guidelines proposing that display density be kept as low as possible, but still retaining the relevant information, are clearly supported by the data. Assuming the information necessary for the task is displayed, there can be little doubt that increasing the number of items beyond that level will have a detrimental effect on performance.

However, the recommendations for specific values of display density are not so clearly supported (e.g., Danchak's, 1976, recommendation that display density not exceed 25%). The main problem is that it is not always possible to convert the measures of density used in the various studies to a common measure for comparison purposes. The different studies used their own unique operational definitions of density -- such as total number of characters, number of rows or columns in an array, number of lines in a table, or number of "facts". For most of these studies it is not possible to derive a percentage of the sort discussed by Danchak (1976). One study that did manipulate density as a percentage of spaces
available (Dodson & Shields, 1978) started at 30%, which is above Danchak's recommended maximum of 25%. They found that performance deteriorated with increasing density; however, the effect of densities under 30% was not studied. One would expect, based on Danchak's recommendation, that increases in density up to about 25% would have little effect on performance while further increases would have a more substantial impact.

In one other study it is possible to convert the measure of density used to a percentage. Vitz (1966) asked subjects to rank order their preferences for random line drawings that varied in number of lines from 4 to 128. Making certain assumptions about the construction of these lines, the extremes of 4 lines and 128 lines correspond to densities of 1.7% and 54%. Using that measure, Vitz's data show that subjects' preferences for the displays increased up to 13.5% density, then decreased from there. Although one cannot draw conclusions about performance based on these subjective ratings, it is interesting that this peak of 13.5% is quite close to Danchak's (1976) recommended optimum density of 15%.

As stated earlier, Danchak (1976, p. 33) reported that his analysis of "good" CRT displays indicated an average density of

---

4. Vitz (1966) constructed the line drawings by randomly selecting a starting point on one side of a square, each side of which was divided into 160 small sections. A random angle between 1 and 179 degrees was then chosen and the line drawn from the starting point at the angle until it intersected another side. Measurements of the 32 lines in Vitz's sample line drawing (his Figure 5) showed that the average line length was 108 units, based upon the length of a side being 160 units. Densities can then be calculated by assuming that each line takes up 108 spaces in the 160 x 160 space array.
15%. To corroborate that figure, a similar analysis of CRT displays from the ARSB system mentioned earlier was conducted. Although these displays have not necessarily been judged as "good," they are displays from a successful, currently operational system. A randomly selected sample of 20 different CRT displays from this system had a mean density of 14.2%, standard deviation of 7.1%, and range of 0.9% to 27.9%. The close agreement between this mean of 14.2% and Danchak's 15% is surprising considering the different populations of displays involved. In addition, the fact that only 2 of the 20 ARSB displays exceeded 25% density (i.e., 25.3% and 27.9%) lends credence to Danchak's recommendation that display density not exceed 25%.

In summary, it is clear that display density should be kept low while still ensuring that the required information is displayed. There is also some indirect support for Danchak's (1976) recommendations of 15% optimal density and 25% maximum density, in that 15% appears to be a common level of density and 25% appears to be a common upper limit. Additional research manipulating density over a wide range, including the lower levels, is needed to address this question directly.

3.3 Local Density

3.3.1 Guidelines

Some of the guidelines mention that the display should contain ample "blank spaces" between the items. This concept is closely related to overall density but is not identical to it. For example, it is possible to have two displays with the same overall
density, but in one case the items could be packed into one corner of the display while in the other case the items could be evenly dispersed across the entire display. This latter case would provide more blank spaces between the items. This concept of how "tightly packed" the items are will be called "local density."

Some guidelines have mentioned that such use of blank spaces can aid in structuring the display: "Blank spaces can...be used to provide structure in the display" (Cakir et al., 1980, p. 114); "Space makes it easier to find one's way around by breaking up the text into logical segments" (Jones, 1978, p. 159); "Blanks help to spatially organize the material" (Peterson, 1979, p. 21); "Spacing and blanks in a display...are important, both to emphasize and maintain the logical sequencing or structure" (Stewart, 1976, p. 142).

In spite of these admonitions that spacing is important in the design of a display, none of the guidelines has offered a general way of quantifying its use (analogous to the overall density percentage). Several of the guidelines have, however, offered the specific recommendation that groups of items be separated by three to five rows or columns of blank spaces (Bailey, 1982, p. 347; Engel & Granda, 1975, p. 8; Galitz, 1980, p. 107).

3.3.2 Empirical Data

Several experiments have studied local density by manipulating the spacing between lines of a display. Ringel and Hammer (1964) manipulated the ratio of letter height to space between lines in a table from 1:4 for low density to 1:2 for high density. They found that the time to answer a question about the display decreased
slightly, but significantly, with greater density of lines. Thus, higher local density enhanced performance. On the other hand, Kolers et al. (1981), in comparing single-spaced vs double-spaced CRT displays of text, found that single spacing required more eye fixations per line, resulted in fewer words read per fixation, and required longer total reading time. Thus, higher local density degraded performance.

The apparent conflict between these two studies can be clarified by a closer examination of the stimuli. As stated, Ringel and Hammer (1964) varied the ratio of letter height to space between lines from 1:2 to 1:4. Although Kolers et al. (1981) did not provide this measure for their stimuli, measurements made from their photographs of the stimuli (their Figure 2, pp. 520-521) reveal that single-spacing was a ratio of about 1:0.3 (i.e., the bottom of one line almost touched the top of the next) and double-spacing was a ratio of about 1:1.7, approximating Ringel and Hammer's 1:2. Thus, the two studies revealed that double-spacing (approximately 1:2) yields optimal performance, while performance degrades on either side of that point.

Several visual search studies have manipulated what may be called local density using a variety of other techniques. Brown and Monk (1975) manipulated the number of background dots in the 3 x 3 matrix of character spaces centered around a target "double dot." They found that search time increased as the number of local background dots increased from 0 to 8. Egeth et al. (1973) manipulated local density by using either a circular or a linear array of 1-5 characters. In the circular array the minimum separation
between characters was 1.37 deg, while in the linear array it was 0.53 deg. Thus, local density was higher for the linear array. They found that time to detect the target character increased more rapidly with increasing total number of characters when the linear array was used than when the circular array was used. Thus, higher local density caused a slower rate of searching.

Treisman (1982), in her Experiment 3, manipulated local density by varying the distance between groups of characters. She found that the search time to find a target character was slightly, but significantly, longer with the less dense displays. Thus, contrary to the other search studies (but consistent with Ringel & Hammer, 1964), she found that higher local density enhanced performance.

Here again the conflict between Treisman's (1982) findings and those of the other search studies (Brown & Monk, 1975; Egeth et al., 1973) can be clarified somewhat by a consideration of the stimuli. In the Brown and Monk (1975) study, the target was separated from the background dots by only 0.21 deg. In the Egeth et al. (1973) study, the target was separated from the other characters by only 0.53 deg or 1.37 deg. On the other hand, Treisman (1982) manipulated the separation of groups of characters from the fixation point over a range of 2.14 deg to 4.28 deg. Thus, Treisman studied much lower local densities than did the others. These findings seem to indicate that at low levels of local density raising the density enhances performance (i.e., Treisman, 1982, Ringel & Hammer, 1964), while at high levels of local density raising the density degrades performance (i.e., Kolers et al., 1981; Brown & Monk, 1975; Egeth et al., 1973). Perhaps such a function could be
explained by lateral masking of characters at high local densities (see Bouma, 1970; Collins & Eriksen, 1967) and by increased eye-movement time at low local densities.

3.3.3 Synthesis

The relationship between the guidelines related to local density and the empirical data is difficult to assess, primarily because the guidelines are poorly defined, stating only that spacing helps to structure a screen. Certainly, none of the guidelines suggests what the empirical data seem to suggest: that there may be an optimal level of local density, below or above which performance degrades.

What is lacking in both the guidelines and the empirical studies is a definition of local density applicable to any alphanumeric display. Conceptually, the term is used to mean how "tightly packed" the display is, but a variety of measures have been used (e.g., line spacing, separation of adjacent characters, separation of groups). Most of these measures are only applicable to particular types of displays. However, the measure used by Brown and Monk (1975) shows promise for being generalizable. They manipulated local density by varying the number of background dots in the 3 x 3 matrix of spaces surrounding the target. In essence, this is a measure of how many background symbols there are near the target. Obviously this could be converted to a percentage, since there are eight character spaces in the surrounding 3 x 3 matrix. To generalize this measure to other tasks besides visual search (i.e., tasks where there is not a specific target character), one could
calculate an average percentage of filled spaces in the 3 x 3 matrix surrounding every character on the display. This would then be an index of local density for an entire display.

This measure of local density has an arbitrary component in the selection of a 3 x 3 matrix surrounding each character. That selection of the surrounding matrix defines what is meant by "near" a character. One would like to make that selection based upon some characteristic of the human visual system. A somewhat less arbitrary choice would be to define the area "near" each character as being the area surrounding the character to which the eye is most sensitive. It is well known that visual acuity falls off rapidly as the visual angle from the point of fixation increases (e.g., Anstis, 1974; Blackwell & Moldauer, 1958; Taylor, 1961). In fact, averaging the data of Blackwell and Moldauer (1958) and Taylor (1961) shows that relative visual acuity has approximately halved at a distance of 2.5 deg from the point of fixation. This is consistent with the data of Bouma (1970), which showed that when a letter was flanked on either side by other letters and presented at 2.5 deg, it was reported with half the accuracy of a letter at the fixation point. Thus, a 5-deg (diameter) circle centered around the character may be a reasonable choice for the area "near" that character. This choice of a 5-deg circle is consistent with Dan- chak, 1976, who selected a 5-deg circle as the basis for calculating the recommended maximum length for a displayed word.

If one assumes that the average CRT-viewing distance is 475 mm (which is the middle of the 450-500 mm "optimum" viewing range stated by Cakir et al., 1980, p.173), then the 5-deg visual angle
translates to a circle with 41.6 mm diameter on the face of a CRT. Likewise, if one assumes that the average distances between character centers are 2.8 mm horizontally and 5.6 mm vertically (based on the average dimensions of characters displayed on three different vendors' CRT's), then the 5-deg visual angle includes about 88 spaces centered around the point of fixation. This is illustrated in Figure 3.

A possible index for local density, then, could be the average percentage of characters in these 88 spaces surrounding each character on the screen. Such an index, however, would fail to account for the different sensitivities of the eye within the 5-deg visual angle. The effect of characters closer to the center of the 5-deg circle is clearly greater than the effect of those at the periphery, due to different sensitivities of the eye and lateral masking. Thus, a more realistic index of local density would weight those characters closer to the center more heavily than those farther out. For ease of calculation, a linear weighting scheme could be chosen, whereby the weight assigned to each character space is inversely proportional to its distance from the fixation point. Such a scheme of weights is illustrated in Figure 3.5 Arbitrarily, the center (fixation) character has been assigned a value of ten,

5. This particular weighting scheme was chosen for computational simplicity. A psychophysically determined scheme would certainly be more appropriate than a simple linear scheme. It would also be appropriate to extend the weights beyond the somewhat arbitrary 5-deg boundary chosen here. The computational complexity of such a scheme, however, is significantly greater than that of the scheme outlined here.
Figure 3. Character spaces on CRT screen subtended by 5-deg visual angle and their approximate weights.
and those character spaces outside the 5-deg circle have been assigned a value of zero.

The proposed index of local density would be the sum of the weights assigned to filled character spaces around each character, expressed as a percentage of the total possible weight. Conceptually, this index can be viewed as the average percentage of other characters near each character, with those closer being weighted more heavily.

A better understanding of this index and its relationship to overall density can be gained by studying Figures 4-7. These are examples of various displays that illustrate how local density and overall density covary. Figure 4 shows the simplest case: an empty display. Obviously, both overall density and local density are 0%. Figure 5 shows the opposite case: a completely filled display. Overall density, obviously, is 100%. Local density is also high, 81%, because each character is surrounded by many others. Local density is not 100% because the characters at the edges of the display are viewed as being surrounded by blank spaces outside the display. Local density would approach 100% as the display gets very large.

Between these limiting cases of an empty and full display, overall density and local density covary in a variety at ways. Figure 6 shows a display in which only the left half is filled, thus halving the overall density to 50%. On the other hand, the local density, 72%, is still high because the characters are densely packed. Figure 7 shows a display that also has only half of the character spaces filled, but they are distributed uniformly
Overall density = 0%
Local density = 0%

Figure 4. An empty display

Overall density = 100%
Local density = 81%

Figure 5. A completely filled display
Figure 6. A display with the left half filled

Figure 7. A display with half the spaces filled uniformly across the display
across the entire display. Obviously, the overall density is still 50%, but the local density has now been reduced to 39%.

It is also instructive to apply this index of local density to real-world displays, such as those studied by Tullis (1981), shown in Figures 1 and 2. The overall densities are 17.9% for the "narrative" format of Figure 1 and 10.8% for the "structured" format of Figure 2. The local densities, on the other hand, are 58.0% for the "narrative" format and 35.6% for the "structured" format. This difference in local density between the "narrative" and "structured" formats could reflect at least one underlying reason for the subjects' better performance with the "structured" format.

In summary, a technique for measuring local density that can be applied to any alphanumeric display has been proposed. While it may be sensitive to aspects of display design that have an effect on human performance, that remains to be shown empirically.

3.4 Grouping

3.4.1 Guidelines

Many of the guidelines recommend that similar items on the display be distinctly grouped. This is closely related to local density, since the primary technique used for grouping items is to leave blank spaces between groups. Most of the guidelines claim that grouping enhances the structure of the display: "Grouping techniques are important in helping to organize information" (Bailey, 1982, p. 330); "Grouping similar items together in a display format improves their readability and can highlight relationships between different groups of data" (Cakir et al., 1980, p. 114);
"Screens should provide cohesive groupings of screen elements so that people perceive large screens to have identifiable pieces" (Galitz, 1980, p. 108); "Information shall be placed in groups to permit the operator to associate or compare like classes of information" (MIL-STD-1472C, 1981, p. 246).

Two of the references take this guideline a step further and propose in general terms how the grouping should be done. Both Cropper and Evans (1968) and Danchak (1976) state that the 5-deg visual angle to which the eye is most sensitive (called the "span of attention" by Danchak) defines discrete areas on the screen and that the display should be structured with this in mind. "The presentation of information in 'chunks'...which can be taken in at one fixation will help to overcome the limitations in the human input system in searching tasks" (Cropper & Evans, 1968, p. 96). Likewise, "since the span of attention defines discrete areas of the CRT, it is advisable to have each area contain only one piece of information" (Danchak, 1976, p. 34).

3.4.2 Empirical Data

The empirical evidence directly relevant to grouping on visual displays is somewhat sparse. Although grouping effects on short-term memory have been studied extensively (e.g., Kahneman & Henik, 1977; Mayzner & Gabriel, 1963; Severin & Rigby, 1963; Winzenz, 1972), such studies are not really relevant to the tasks involved in the use of complex real-world displays.

Several studies have investigated the effects of different semantic bases for groupings on real-world displays. For example,
Dodson and Shields (1978) found that the time to locate an item on a display was shorter when the items were grouped by function (i.e., by instrument involved) than when the groups had no functional basis. Likewise, Card (1982), in his Experiment 1, found that the time to select an item from a menu varied with the semantic basis of the groups: alphabetical grouping was fastest, functional grouping was next, and random grouping was slowest. In his Experiment 3, Card found that subjects tended to recall menu items in "chunks" which corresponded to the menu groups. Chunk boundaries were defined by a 2 sec or more pause during recall, analogous to the technique used in studies of chunking in chess perception by Chase and Simon (1973). Although these studies are informative, they do not address the primary topic of this discussion, which is the spatial aspect of grouping (e.g., grouping vs lack of grouping) as opposed to the semantic aspect.

A few studies have manipulated the spatial component of grouping. Woodward (1972) found that the time to compare three-digit numbers in pairs was shorter when the numbers were presented side-by-side (proximal) than when they were presented end-to-end (distal). As Woodward points out (p. 338), both numbers fell within a 5-deg visual angle in the proximal arrangement (thus forming a "group") while only one number fell within that space in the distal arrangement.

Banks and Prinzmetal (1976) found that the time to detect a target letter was shorter when the letter was in a group by itself than when it was part of a group of noise characters. A group was defined as being an arrangement of characters in "good form"
according to the definition used by Garner (1970, 1974): a good form generates a small number of different forms when reflected about an axis or rotated 90 degrees (i.e., it has symmetry). Banks and Prinzmetal's finding cannot be attributed to target-noise proximity (i.e., local density for the target) because in the "grouped with noise" conditions, the target was just as far from the noise characters as it was when grouped separately.

Banks and Prinzmetal also provided an interesting test of their assumptions about the subjects' perception of groups (i.e., that characters in good form are perceived as a group). They simply asked subjects to draw boundaries around items that seemed to group together on the displays. They derived a measure of target-noise grouping by counting the number of boundaries drawn between the target and noise characters and then dividing this by the number of noise characters. The higher this number was, the more isolated the target was. As expected, they found that this grouping index was twice as large for the condition where they intended the target to be perceived as being separate than the condition where the target was part of a good form with noise characters.

Treisman (1982) provided one of the most extensive studies of grouping. In her Experiment 1 she manipulated grouping by displaying 36 characters in either 1, 4, 9, 18, or 36 groups. Each group was composed of homogeneous letters (i.e., all green X's or red H's), except for the group containing the target (i.e., a green H). She found that the time to detect a target letter requiring a conjunction of features (i.e., color and letter shape) increased as the number of groups increased, even though the total number of
characters remained the same. The function relating reaction time to number of groups was negatively accelerated, with increases in number of groups at the low end (i.e., 1-9 groups) having the largest effect on reaction time. On the other hand, she found that number of groups had little effect on time to detect a target letter defined by a disjunctive feature (e.g., a blue H among red H's and green X's). This is consistent with her earlier finding (Treisman & Gelade, 1980) that visual search for targets defined by one or more disjunctive features occurs in parallel.

The conclusion to be drawn from these various studies of grouping is not particularly clear. Woodward (1972) found that grouping numbers together enhanced the ability to compare them. Banks and Prinzmetal (1976) found that grouping a target separately from the noise enhanced the ability to detect the target. Treisman (1982) found that grouping characters into larger groups enhanced the ability to detect a conjunctively defined target. The problem is that one cannot really make comparisons between these studies because of the differences in the tasks and the differences in the types of grouping. Each of the studies used its own operational definition of grouping: Woodward (1972) viewed a group as being two 3-digit numbers displayed proximally to each other; Banks and Prinzmetal (1976) viewed a group as being an arrangement of characters having "good form"; Treisman (1982) viewed a group as being a set of adjacent, homogeneous characters. In addition, only Banks and Prinzmetal (1976) took the step to show, empirically, that their operationally defined groups correspond to groups perceived by the subjects.
3.4.3 **Synthesis**

The empirical data seem to support the idea that grouping of items is beneficial to performance. A smaller number of groups is better than a larger number of individual items (Treisman, 1982). In another sense, grouping is beneficial if the key item can be grouped by itself (Banks and Prinzmetal, 1976). However, neither the guidelines nor the empirical studies have presented a general definition of grouping. Perhaps the closest to that is the definition used by Banks and Prinzmetal (1976), derived from Garner (1970, 1974): a group is a collection of symbols constituting a "good form." Such a definition, however, seems limited in its generalizability to complex real-world displays composed of words and alphanumeric data with a variety of spatial relationships.

Another type of definition of grouping was proposed by Cropper and Evans (1968) and Danchak (1976), who proposed that a screen should be designed in discrete "chunks," each of which subtends a visual angle of less than 5 deg. Precisely how such a definition could be applied, however, is not clear. For example, the definition is circular, in that it invokes the concept of a "discrete chunk" (which is apparently the same thing as a group) in defining what it is that must fall within a 5-deg visual angle. Obviously, it is not enough to say that anything that falls within a 5-deg visual angle is a group, since every character on any display would fall within a 5-deg visual angle centered on that character. The key is the term "discrete," which implies that the character, or set of characters, has a boundary that separates it from other characters. Perhaps that boundary is defined by blank spaces (but
how many?), by some special characters (e.g., asterisks), by a change from upper- to lower-case, or by some other technique altogether.

This question of what defines a "group" is certainly not a new one in psychology. The Gestalt psychologists were the first to study the problem extensively. They attempted to understand perception in terms of a series of organizational laws or principles, including the laws of proximity, similarity, continuity, and common fate (Wertheimer, 1923; Koffka, 1935).

Some attempts have been made to expand the Gestalt laws and implement them in computer programs for detecting groups. For example, Zahn (1971) described a technique for implementing the law of proximity. Although he used a complex technique for detecting groups of dots, the following simplification may be sufficient for detecting groups of alphanumeric items. Consider the display shown in Figure 8. First, compute the distance between each character and its nearest neighbor. Using on-center distances between characters of 1 unit horizontally and 2 units vertically, this results in a mean of 1.42. Then form a graph by connecting any pair of characters whose distance is smaller than a threshold value depending on that mean. A threshold value of twice that mean seems to

7. Zahn (1971) actually used a threshold value based on the mean of the distances between each dot and its nearest neighbor as well as the standard deviation of those distances. However, the standard deviation is of little use with alphanumeric displays since it tends to be very close to zero. That is, there are relatively few isolated characters on normal CRT displays; most are contiguous with at least one other character.
Figure 8. Sample display to illustrate grouping algorithm

Figure 9. Sample display with lines connecting all pairs of characters separated by less than a threshold distance
work well for alphanumeric CRT displays. Such a graph is shown in Figure 9. A group is then defined as any interconnected set of characters. Thus, Figure 9 represents two groups. Although this technique seems to identify groups that match with intuition, Zahn (1971) does not give any empirical data to demonstrate its validity.

Some insights can be gained into this technique for identifying groups by applying it to real-world CRT displays. Once again, the displays from the Tullis (1981) study shown in Figures 1 and 2 will be used. The "narrative" format of Figure 1 has a mean distance between neighboring characters of 1.05 (using the same units as in the example of Figures 8 and 9). The "structured" format of Figure 2 has a mean distance between neighboring characters of 1.09. Connecting those pairs of characters separated by less than twice the mean (that is, 2.10 for Figure 1 and 2.18 for Figure 2) results in the configurations shown in Figures 10 and 11.

Figure 10 shows that the "narrative" format contains only three groups of connected characters. On the other hand, Figure 11 shows that the "structured" format contains thirteen groups of connected characters. The design of the "structured" format, then, is more consistent with the various guidelines stating that a large screen should be broken into a number of small, discrete groups. Perhaps this difference in the number of groups could be another factor mediating the subjects' better performance with the "structured" format.

The obvious question this raises is, "What is the optimum number of groups?", if such a number exists. Treisman's (1982) studies
Figure 10. Narrative format from Tullis (1981) with grouping indicated

Figure 11. Structured format from Tullis (1981) with grouping indicated
would seem to suggest that the fewer groups the better. One should be careful, however, in making such a generalization from her studies. For example, her largest groups were composed of only 36 characters, unlike the 319 characters in the largest group of Figure 10. Perhaps if she had manipulated grouping over a wider range she would have found that when using a few large groups, dividing the information into smaller groups might enhance performance; but when using many small groups, dividing the information into even more groups might degrade performance.

Another aspect of the groupings proposed by Cropper and Evans (1968) and Danchak (1976) is the size of the groups. It is possible to measure the visual angle subtended by each of the groups on a display and then derive an average by weighting each group’s visual angle by the number of characters in the group. Applying this technique to the "narrative" format of Figure 10 results in an average visual angle of 13.3 deg. Likewise, the "structured" format Figure 11 results in 5.2 deg. Obviously, the "structured" format more closely adheres to the recommendations (Cropper & Evans, 1968; Danchak, 1976) that a group should fit within a 5-deg visual angle. Perhaps this difference could be another factor mediating the difference in subjects' performance with these two formats. Possibly this percentage measure could be a more predictive index of grouping than simply the absolute number of groups.
3.5 Layout Complexity

3.5.1 Guidelines

One of the most common themes in the guidelines, and perhaps the most difficult to characterize, is that the overall display format should minimize the complexity of the layout, or maximize the visual predictability. That is, based on a knowledge of where some items appear on the screen, one should be able to predict the locations of others. This concept is reflected by the guidelines in a variety of ways: "People seem to be able to scan for a certain item of information more quickly and accurately if a tabular format...is provided" (Bailey, 1982, p. 345); "Use vertically aligned lists with left justification for most rapid scanning" (Engel & Granda, 1976, p. 6); "An almost visual motion can be created by either implicit or explicit 'lines' formed by the display elements" (Green, 1976, p. 147); "Tabular formats of alphanumeric data have been shown to be most usable and accurately readable" (NASA, 1980, p. 3-20); "People can scan for information most accurately if columnar formats are used" (Peterson, 1979, p. 34).

In addition to these general recommendations about using tables, columnar lists, and vertical alignments, a number of the guidelines make the specific recommendation that words and alphanumeric data should be left-justified while numeric data should be right-justified on the decimal point (e.g., Bailey, 1982, p. 346; Engel & Granda, 1976, p. 7; Galitz, 1980, p. 113).

The effect of all these recommendations is to increase the predictability of the locations of items on the display (or, with
numbers, the predictability of the decimal point's location. For example, in a tabular format, the location of any entry in the table can be predicted from a knowledge of the column and row positions. By contrast, the location of any item in a textual format cannot be predicted so efficiently.

Marcus (1981, 1982) has incorporated this concept of minimizing layout complexity into a general system for the design of an interface. He proposes a "grid system" which uses a few imaginary vertical and horizontal lines to determine the positioning of elements on the display. In essence, this approach limits the amount of spatial variation between display elements and enhances visual predictability.

3.5.2 Empirical Data

Despite the number of guidelines addressing this topic, there have been virtually no studies directly addressing it. The only study that comes close is that by Brown and Monk (1975). They used two different patterns of background dots in a visual search task: random or constrained. This was manipulated by the probability, P, of transition from "dot" to "no dot" in adjacent positions of the background matrix. In the random condition, P = 0.5 (i.e., the transition from dot to no dot was just as likely as the opposite transition). In the constrained condition, P = 0.25 (i.e., the transition from dot to no dot was less likely than the opposite transition-- so that once the condition "dot" was established in the pattern, it was likely to continue for a while). This manipulation resulted in a less complex display for the constrained con-
dition. They found that search time to find a target "double dot" was greater with the random background than with the constrained background. (This effect was independent of local density of background dots surrounding the target.) In essence, then, the constrained background enhanced the search process.

Although this study is interesting, it is only marginally relevant. The manipulation of random vs constrained background does not address exactly the same kind of layout complexity as described in the guidelines. The guidelines focus on reducing layout complexity by aligning data items vertically and horizontally, while Brown and Monk (1975) used a probabilistic approach to reducing layout complexity.

3.5.3 Synthesis

Although Brown and Monk's (1975) study lends some support to the guidelines addressing layout complexity, the support is weak at best. Perhaps what is needed to stimulate research in this area is a general technique for measuring layout complexity.

Bonsiepe (1968) proposed a method for quantifying the layout complexity of a typographically-designed page that he used to compare before and after versions of a page from a catalog, shown in Figures 12 and 13. The technique involves first drawing rectangles around all the groups of items on the page. These rectangles should correspond to the groups' "real or maximum possible extents" (p. 209) and should not overlap with other groups. Figures 14 and 15 illustrate the rectangles Bonsiepe drew on the old and new catalog pages.
Figure 12. Original catalog page from Bonsiepe (1968)

Figure 13. Redesigned catalog page from Bonsiepe (1968)
Figure 14. Original catalog page from Bonsiepe (1968) showing rectangles defining each group

Figure 15. Redesigned catalog page from Bonsiepe (1968) showing rectangles defining each group
Bonsiepe (1968) derived measures of two types of order from these rectangles: system order and distribution order. The "system order" is determined by counting the number of unique widths and heights for the rectangles on a page. The old version (Figure 14) has 19 widths divisible into 9 classes and 19 heights divisible into 14 classes. The new version (Figure 15) has 20 widths divisible into only 3 classes and 20 heights divisible into only 5 classes. Bonsiepe then used the following formula adapted from information theory (Shannon & Weaver, 1949) to calculate the complexity of this system:

\[ C = -N \sum_{n=1}^{m} p_n \log_2 p_n \]

where:

- \( C \) = Complexity of the system, expressed in bits
- \( N \) = Number of events (i.e., widths or heights)
- \( m \) = Number of event classes (i.e., no. of unique widths or heights)
- \( p_n \) = Probability of occurrence of the nth event class (based on the frequency of events within that class)

Applying this formula to the old design results in 53 bits being conveyed by the widths and 70 bits by the heights. Assuming that the system order results from both of these factors, they may be added to give a total of 123 bits for the old design. Similar calculations for the new design result in 18 bits for widths and 37
bits for heights, giving a total of 55 bits. Thus, using this measurement technique, the new version is about 55% simpler than the old.

The second type of order, distribution order, involves the layout of the groups of items on the page. This layout can be specified by measuring the horizontal and vertical distances of each group (or, actually, its rectangle) from some starting point on the page. Arbitrarily, Bonsiepe chose to measure the horizontal distance of the rectangle’s left side from the left edge of the page, and the vertical distance of the rectangle’s top side from the top edge of the page. The following measurements were then obtained: for the old version, 19 vertical distances divisible into 17 classes, and 19 horizontal distances divisible into 6 classes; for the new version, 20 vertical distances divisible into 10 classes and 20 horizontal distances divisible into 4 classes. Using the measure of complexity results in the following:

Old version: 90 bits for vertical distances
+ 37 bits for horizontal distances
—
127 bits

New version: 63 bits for vertical distances
+ 35 bits for horizontal distances
—
98 bits

Thus, the distribution order of the new version is about 23% simpler than that for the old.

Bonsiepe’s technique has some intuitive appeal, in that the new
version of the catalog page certainly appears to be more orderly than the old version. He does not, however, present any empirical evidence to demonstrate the validity of the technique.

In attempting to apply Bonsiepe's technique for measuring layout complexity to CRT displays, one encounters some problems. The major problem is deciding what to draw the rectangles around. An examination of Figures 12 through 15 reveals that Bonsiepe's rectangles were somewhat subjectively determined. For example, in the old version shown in Figures 12 and 14, why is the word "Gegenrahmen" under the first picture grouped with the following text instead of by itself? Likewise, how does one determine the "real or maximum possible extent" of each group?

These problems with Bonsiepe's technique led to the adoption of a modified technique for application to CRT displays. First, the concept of "system order," or the order that is dependent on the size of each group, was dropped. This was done primarily because of the ill-defined nature of a group's "maximum possible extent." In addition, it appears that the other type of order, distribution order, is more appropriate to CRT displays because of the tasks they are commonly used for. Since many of the tasks involve searching for and then encoding items on the display, it would appear that the predictability of the beginning of each item (i.e., where one would usually want to start encoding) is more important than the width and height of the item. This brings up the second modification, involving what items on the display to consider in measuring the distribution order. Formatted CRT displays almost always contain two types of items: labels and data. The labels
remain basically the same on different representations of the same display while the data change. A single label as well as a single data item may be composed of more than one "word" (group of contiguous characters). These labels and data items are the units upon which the distribution order will be measured, since they appear to be the units a user would search for and encode.

To illustrate this modified measure of the layout complexity of a CRT display, consider the displays shown in Figures 16 and 17, taken from Stewart (1976). Stewart presented Figure 17 as a redesigned version of Figure 16. Although he did not discuss layout complexity per se, his redesign presents a striking example of the use of techniques to maximize distribution order. The labels and data items on both displays have been inscribed by rectangles. The same basic items are inscribed on both figures, although Stewart occasionally changed the wording. Also, the redesigned version has fewer labels and data items. An analysis of the distribution order results in the following:

**Figure 16 - Original Version:**

36 horizontal distances in 22 unique classes = 140 bits  
36 vertical distances in 11 unique classes = 122 bits  

Overall complexity = 262 bits

**Figure 17 - Redesigned Version:**

30 horizontal distances in 6 unique classes = 67 bits  
30 vertical distances in 10 unique classes = 98 bits  

Overall complexity = 165 bits
Figure 16. Original version of display from Stewart (1976) with rectangles added to indicate labels and data items

Figure 17. Redesigned version of display from Stewart (1976) with rectangles added to indicate labels and data items
Thus, the distribution order of the redesigned version is about 37% simpler than the original version. Most of the improvement in order resulted from the adoption of a small number of horizontal positions (i.e., "tab stops") for the beginning of data items, thus reducing the horizontal complexity by more than half.

Regrettably, Stewart (1976) does not present any empirical evidence related to the use of these displays. Thus, the validity of the layout complexity measure cannot be addressed.

Since the measurement techniques described for all the other characteristics (overall density, local density, grouping) have been applied to the displays from Tullis (1981), this measure of layout complexity will be applied to them as well. Figures 18 and 19 show the "narrative" and "structured" formats from that study, with the labels and data items inscribed by rectangles. The same basic items are inscribed in both cases, although the "narrative" format has more labels. Note that the technique used for inscribing numeric data was based on the maximum values that the data could assume.

The results of applying Bonsiepe's (1968) technique for measuring distribution order are as follows:

Figure 18 - Narrative Format:

22 horizontal distances in 6 unique classes = 41 bits
22 vertical distances in 20 unique classes = 93 bits

Overall complexity = 134 bits
**Figure 18.** Narrative format from Tullis (1981) with labels and data items inscribed by rectangles

**Figure 19.** Structured format from Tullis (1981) with labels and data items inscribed by rectangles
Figure 19 - Structured Format:

18 horizontal distances in 7 unique classes = 43 bits
18 vertical distances in 8 unique classes = 53 bits

Overall Complexity = 96 bits

Thus, the distribution order of the structured format is about 28% simpler than the narrative format. The improvement in layout complexity resulted from the lower vertical complexity of the structured format's tabular layout compared to the narrative format's list layout. Here again, perhaps this difference could be another factor mediating the subjects' better performance with the structured format.

3.6 Summary

Four characteristics of alphanumerical display formats have been discussed: overall density, local density, grouping, and layout complexity. For the first three of these there is some empirical evidence indicating that the characteristic has an effect on human performance. For the last characteristic discussed, layout complexity, there is virtually no empirical evidence directly addressing its effect on human performance. Numerous guidelines suggest, however, that increasing the layout complexity may have a detrimental effect on human performance.

Much of the discussion has focused on deriving objective ways of measuring these four characteristics. As a summary of these measures, Table 3 presents a definition of each and shows the results of applying them to the "narrative" and "structured" displays (Fig-

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Definition of Measure</th>
<th>Narrative Display</th>
<th>Structured Display</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Density</td>
<td>Number of filled character spaces as a percentage of total spaces available</td>
<td>17.9%</td>
<td>10.8%</td>
</tr>
<tr>
<td>Local Density</td>
<td>Average number of filled character spaces in a 5-deg visual angle around each character, expressed as a percentage of available spaces in the circle, and weighted by distance from the character</td>
<td>58.0%</td>
<td>35.6%</td>
</tr>
<tr>
<td>Grouping</td>
<td>(1) Number of groups of &quot;connected&quot; characters, where a connection is any pair of characters separated by less than twice the mean of the distances between each character and its nearest neighbor (2) Average visual angle subtended by groups (as defined above), weighted by number of characters in the group</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>Layout</td>
<td>The complexity, as defined in information theory, of the distribution of horizontal and vertical distances of each label and data item from a standard point on the display</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complexity</td>
<td></td>
<td>134 bits</td>
<td>96 bits</td>
</tr>
</tbody>
</table>
ures 1 and 2) from the Tullis (1981) study.

All of the measures showed a difference between the "narrative" and "structured" formats. The differences indicate that these measures, or some combination of them, could provide the basis for objectively evaluating the usability of a display, since Tullis (1981) did find performance differences between the two formats. The predictive validity of these measures, however, cannot be assessed from the Tullis (1981) study since the characteristics were confounded with each other. The differences in performance could have arisen from the differences in some of the characteristics, all of the characteristics, or even some other characteristics not currently defined. A study manipulating these characteristics in a controlled fashion needs to be conducted. Such a study could provide the basis for a "display evaluation tool" that could be used to predict the usability of a display without the cost of collecting performance data.
4. EXPERIMENT 1: DEVELOPMENT OF THE PREDICTION SYSTEM

4.1 Introduction

The next step in developing a tool for predicting the usability of alphanumeric displays was to develop a computer program for measuring the display characteristics described earlier. The necessity of using a computer to measure these characteristics should be obvious from the complexity of the techniques, especially those for measuring local density, grouping, and layout complexity.

A computer program for measuring the display characteristics was written in the "C" programming language to run under the UNIX® operating system. A listing of the program is included in Appendix A. In order to run the program the following inputs must be provided:

1. A literal example of the display to be analyzed, stored in a UNIX file

2. Characteristics of the device on which the display will be shown:
   - Maximum number of rows and columns of characters (e.g., 24 x 80)
   - Distances between character centers vertically and horizontally

---

8. UNIX is a trademark of Bell Laboratories
3. Viewing distance of the display from the user (needed for calculating visual angles)

A sample output from the program is shown in Figure 20 for one of the displays used in Experiment 1. The output contains the following:

1. The parameters input to the program (i.e., characteristics of the display device, viewing distance)

2. The display file as it was input to the program

3. The groups of characters on the display detected by the grouping algorithm. (Each group is represented by a different symbol.)

4. Results of six measurements:
   - Overall density
   - Local density
   - Number of groups
   - Average visual angle of the groups
   - Number of items
   - Item uncertainty

All six of these measurements were made using the techniques described earlier. However, the last two measurements (number of items and item uncertainty) are actually the two components that
This analysis is based on the following parameters:

- **Characteristics of display device:**
  - Maximum number of characters per row = 80
  - Maximum number of rows = 24
  - Distances between character centers:
    - Horizontally = 0.100000
    - Vertically = 0.200000
  - Viewing distance = 19.000000

File air1.1:

<table>
<thead>
<tr>
<th>From</th>
<th>Departs</th>
<th>Arrives</th>
<th>Flight</th>
<th>Fares 1st</th>
<th>Fares Coach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asheville, NC</td>
<td>7:20a</td>
<td>8:05a</td>
<td>PI 299</td>
<td>92.57</td>
<td>66.85</td>
</tr>
<tr>
<td></td>
<td>10:10a</td>
<td>10:55a</td>
<td>PI 203</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4:20p</td>
<td>5:00p</td>
<td>PI 239</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Austin, TX</td>
<td>8:15a</td>
<td>11:15a</td>
<td>EA 530</td>
<td>263.00</td>
<td>221.00</td>
</tr>
<tr>
<td></td>
<td>8:40a</td>
<td>11:39a</td>
<td>DL 212</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2:00p</td>
<td>5:00p</td>
<td>DL 348</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7:15p</td>
<td>11:26p</td>
<td>DL 1654</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baltimore, MD</td>
<td>7:00a</td>
<td>8:35a</td>
<td>DL 1767</td>
<td>209.00</td>
<td>167.00</td>
</tr>
<tr>
<td></td>
<td>7:50a</td>
<td>9:32a</td>
<td>EA 631</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8:45a</td>
<td>10:20a</td>
<td>DL 1610</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11:15a</td>
<td>12:35p</td>
<td>EA 147</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1:35p</td>
<td>3:10p</td>
<td>DL 1731</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2:35p</td>
<td>4:16p</td>
<td>EA 141</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

11111111111

3333

44444444 55555555 66666666 22222222

7777777777 77

888888 999999 :: :: :: :: :: :: :: ::

888888 999999 :: :: :: :: :: ::

888888 999999 :: :: :: :: :: ::

--.-- --

>>>>> ??????? @@ @@ AAAAAA BBBBBB

>>>>> ??????? @@ @@

>>>>> ??????? @@ @@

>>>>> ??????? @@ @@

CCCCCCCCCC CC

DDDDDD EEEEEE FF FFFF GGGGGG HHHHHH

DDDDDD EEEEEE FF FFFF

DDDDDD EEEEEE FF FFFF

DDDDDD EEEEEE FF FFFF

DDDDDD EEEEEE FF FFFF

DDDDDD EEEEEE FF FFFF

DDDDDD EEEEEE FF FFFF

<table>
<thead>
<tr>
<th>Filename</th>
<th>Density</th>
<th>Density</th>
<th>#Grps</th>
<th>Angle of Grps</th>
<th>#Items</th>
<th>Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>air1.1</td>
<td>16.9%</td>
<td>39.2%</td>
<td>24</td>
<td>2.9 deg</td>
<td>57</td>
<td>6.71 bits</td>
</tr>
</tbody>
</table>

Figure 20. Sample Output from Program
make up the layout complexity measure described earlier (Section 3.5). Number of items is a measure of how many data items or labels are on the display, where an item is defined as a set of characters on a single display line separated by no more than one blank space. Thus, in Figure 20, the word "Departs" is a single item, as is the phrase "Asheville, NC." Item uncertainty, on the other hand, is the average uncertainty of any one item's location, based on the number of unique positions of items on the display and the frequency of occurrence of each of those positions. As used by Bonsiepe (1968), these two components were multiplied together to yield the total amount of information (in bits) conveyed by the layout of all the items on the display. In the present experiment, however, it seemed more useful to keep the two components separate since it might be possible to determine which component is a better predictor of display usability.

Given these six measures of display characteristics, the purpose of Experiment 1 was to develop a system for predicting the usability of a display based on the measures. Part of the problem was to determine which of the six measures, if any, are predictors of user performance with the displays. The general approach was to develop a wide variety of display formats differing on these measures. Multiple regressions were then used to fit search times and subjective ratings of the displays using the display measures.
4.2 Method

4.2.1 Displays

Displays for presenting two types of data were developed: airline listings and motel listings. These were chosen simply because of their familiarity for most people. Specifically, 26 different formats for presenting airline listings and 26 different formats for presenting motel listings were developed.

Each airline display contained information on flights to a particular city. The flights' originating cities were listed in alphabetical order, and within each city the flights were listed by departure time. Other information included the first class and coach fares, arrival times, and flight numbers. Two examples of the airline displays used are shown in Figure 21.

Each motel display contained listings of motels and hotels for a particular state. Cities were listed alphabetically, and within each city the motels were listed alphabetically by name. Other information included the area code, telephone number, single room rate, and double room rate. Two examples of the motel/hotel displays are shown in Figure 22.

For each of the 52 display formats, 10 examples containing different data were developed. Thus, a total of 520 displays were used in the experiment. In order to illustrate the resulting values for the six display measures, Table 4 shows their means, standard deviations, and ranges.

No attempt was made to develop displays that would provide orthogonal combinations of the six display measures. That would
### To: Atlanta, GA

<table>
<thead>
<tr>
<th>Flight</th>
<th>Depart</th>
<th>Arrive</th>
<th>Flight</th>
<th>Depart</th>
<th>Arrive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asheville, NC</td>
<td>7:20a</td>
<td>8:05a</td>
<td>FI 299</td>
<td>10:10a</td>
<td>10:55a</td>
</tr>
<tr>
<td></td>
<td>4:20p</td>
<td>5:00p</td>
<td>PI 259</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Austin, TX</td>
<td>8:15a</td>
<td>11:15a</td>
<td>EA 530</td>
<td>8:40a</td>
<td>11:39a</td>
</tr>
<tr>
<td></td>
<td>2:00p</td>
<td>5:00p</td>
<td>DL 348</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7:15p</td>
<td>11:26p</td>
<td>DL 1654</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baltimore, MD</td>
<td>7:00a</td>
<td>8:35a</td>
<td>DL 1767</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7:50a</td>
<td>9:32a</td>
<td>EA 631</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8:45a</td>
<td>10:20a</td>
<td>DL 1610</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11:15a</td>
<td>12:35p</td>
<td>EA 147</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1:35p</td>
<td>3:10p</td>
<td>DL 1731</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2:35p</td>
<td>4:16p</td>
<td>EA 141</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### To: Knoxville, TN

<table>
<thead>
<tr>
<th>Flight</th>
<th>Depart</th>
<th>Arrive</th>
<th>Flight</th>
<th>Depart</th>
<th>Arrive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atlanta, GA</td>
<td>9:28a</td>
<td>10:10a</td>
<td>FI: DL 1704</td>
<td>1st:</td>
<td>97.00</td>
</tr>
<tr>
<td></td>
<td>12:28p</td>
<td>1:10p</td>
<td>FI: DL 152</td>
<td>1st:</td>
<td>97.00</td>
</tr>
<tr>
<td>Atlanta, GA</td>
<td></td>
<td>4:50p</td>
<td>FI: DL 418</td>
<td>1st:</td>
<td>97.00</td>
</tr>
<tr>
<td></td>
<td>7:41p</td>
<td>8:25p</td>
<td>FI: DL 1126</td>
<td>1st:</td>
<td>97.00</td>
</tr>
<tr>
<td>Chicago, Ill.</td>
<td>1:45p</td>
<td>5:39p</td>
<td>FI: AL 58</td>
<td>1st:</td>
<td>190.00</td>
</tr>
<tr>
<td>Cincinnati, OH</td>
<td>6:30p</td>
<td>9:35p</td>
<td>FI: DL 675</td>
<td>1st:</td>
<td>190.00</td>
</tr>
<tr>
<td>Dallas, TX</td>
<td>6:50p</td>
<td>9:55p</td>
<td>FI: RC 398</td>
<td>1st:</td>
<td>190.00</td>
</tr>
<tr>
<td>Dayton, OH</td>
<td>5:25p</td>
<td>6:30p</td>
<td>FI: FW 455</td>
<td>1st:</td>
<td>118.00</td>
</tr>
<tr>
<td>Dayton, OH</td>
<td>11:20a</td>
<td>1:10p</td>
<td>FI: FW 453</td>
<td>1st:</td>
<td>189.00</td>
</tr>
<tr>
<td>Detroit, Mich.</td>
<td>9:10a</td>
<td>1:10p</td>
<td>FI: FW 453</td>
<td>1st:</td>
<td>183.00</td>
</tr>
</tbody>
</table>

---

*Figure 21. Two Examples of Airline Displays Used in Experiment 1*
### South Carolina

<table>
<thead>
<tr>
<th>City</th>
<th>Motel/Hotel</th>
<th>Area Code</th>
<th>Phone</th>
<th>Rates</th>
<th>Single</th>
<th>Double</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charleston</td>
<td>Best Western</td>
<td>803</td>
<td>747-0961</td>
<td>$26</td>
<td>$30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Days Inn</td>
<td></td>
<td>881-1800</td>
<td>$18</td>
<td>$24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Holiday Inn N</td>
<td></td>
<td>744-1621</td>
<td>$36</td>
<td>$46</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Holiday Inn SW</td>
<td></td>
<td>556-7100</td>
<td>$33</td>
<td>$47</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Howard Johnsons</td>
<td></td>
<td>524-4140</td>
<td>$31</td>
<td>$36</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ramada Inn</td>
<td></td>
<td>774-8281</td>
<td>$33</td>
<td>$40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sheraton Inn</td>
<td></td>
<td>744-2401</td>
<td>$34</td>
<td>$42</td>
<td></td>
</tr>
<tr>
<td>Columbia</td>
<td>Best Western</td>
<td>803</td>
<td>796-9400</td>
<td>$29</td>
<td>$34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carolina Inn</td>
<td></td>
<td>799-8200</td>
<td>$42</td>
<td>$48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Days Inn</td>
<td></td>
<td>736-0000</td>
<td>$23</td>
<td>$27</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Holiday Inn NW</td>
<td></td>
<td>794-9440</td>
<td>$32</td>
<td>$39</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Howard Johnsons</td>
<td></td>
<td>772-7200</td>
<td>$25</td>
<td>$27</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quality Inn</td>
<td></td>
<td>772-0270</td>
<td>$34</td>
<td>$41</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ramada Inn</td>
<td></td>
<td>796-2700</td>
<td>$36</td>
<td>$44</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vagabond Inn</td>
<td></td>
<td>796-6240</td>
<td>$27</td>
<td>$30</td>
<td></td>
</tr>
</tbody>
</table>

### Virginia

<table>
<thead>
<tr>
<th>City</th>
<th>Motel/Hotel</th>
<th>Area Code</th>
<th>Phone</th>
<th>Rates</th>
<th>Single</th>
<th>Double</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alexandria</td>
<td>Best Western Old Town</td>
<td>703</td>
<td>548-6400</td>
<td>$41</td>
<td>$49</td>
<td></td>
</tr>
<tr>
<td>Alexandria</td>
<td>Days Inn Alexandria</td>
<td>703</td>
<td>354-4500</td>
<td>$32</td>
<td>$36</td>
<td></td>
</tr>
<tr>
<td>Alexandria</td>
<td>Howard Johnsons</td>
<td>703</td>
<td>768-3300</td>
<td>$38</td>
<td>$48</td>
<td></td>
</tr>
<tr>
<td>Alexandria</td>
<td>Imperial 400</td>
<td>703</td>
<td>780-4400</td>
<td>$40</td>
<td>$45</td>
<td></td>
</tr>
<tr>
<td>Alexandria</td>
<td>Mt Vee</td>
<td>703</td>
<td>780-2500</td>
<td>$28</td>
<td>$36</td>
<td></td>
</tr>
<tr>
<td>Alexandria</td>
<td>Ramada Inn</td>
<td>703</td>
<td>751-4510</td>
<td>$55</td>
<td>$59</td>
<td></td>
</tr>
<tr>
<td>Alexandria</td>
<td>Travelers</td>
<td>703</td>
<td>768-2510</td>
<td>$29</td>
<td>$32</td>
<td></td>
</tr>
<tr>
<td>Ashland</td>
<td>Best Western Hanover</td>
<td>804</td>
<td>796-8045</td>
<td>$22</td>
<td>$40</td>
<td></td>
</tr>
<tr>
<td>Ashland</td>
<td>Days Inn</td>
<td>804</td>
<td>796-4262</td>
<td>$24</td>
<td>$28</td>
<td></td>
</tr>
<tr>
<td>Ashland</td>
<td>Econo-Travel</td>
<td>804</td>
<td>798-9221</td>
<td>$23</td>
<td>$27</td>
<td></td>
</tr>
<tr>
<td>Ashland</td>
<td>Holiday Inn</td>
<td>804</td>
<td>798-6231</td>
<td>$23</td>
<td>$29</td>
<td></td>
</tr>
<tr>
<td>Ashland</td>
<td>Kings Motor Inn</td>
<td>804</td>
<td>798-9291</td>
<td>$26</td>
<td>$30</td>
<td></td>
</tr>
<tr>
<td>Bedford</td>
<td>Bedford Inn</td>
<td>703</td>
<td>586-1028</td>
<td>$15</td>
<td>$18</td>
<td></td>
</tr>
<tr>
<td>Bedford</td>
<td>Peaks of Otter Lodge</td>
<td>703</td>
<td>586-1081</td>
<td>$30</td>
<td>$42</td>
<td></td>
</tr>
<tr>
<td>Bedford</td>
<td>Town Terrace</td>
<td>703</td>
<td>586-8286</td>
<td>$21</td>
<td>$25</td>
<td></td>
</tr>
<tr>
<td>Blacksburg</td>
<td>Econo-Travel</td>
<td>703</td>
<td>951-4242</td>
<td>$19</td>
<td>$22</td>
<td></td>
</tr>
<tr>
<td>Blacksburg</td>
<td>Holiday Inn</td>
<td>703</td>
<td>951-1330</td>
<td>$32</td>
<td>$40</td>
<td></td>
</tr>
<tr>
<td>Blacksburg</td>
<td>Marriott Inn</td>
<td>703</td>
<td>552-7001</td>
<td>$34</td>
<td>$42</td>
<td></td>
</tr>
</tbody>
</table>

Figure 22. Two Examples of Motel Displays Used in Experiment 1
<table>
<thead>
<tr>
<th></th>
<th>Standard</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Deviation</td>
</tr>
<tr>
<td>Overall Density</td>
<td>27.9%</td>
<td>12.0%</td>
</tr>
<tr>
<td>Local Density</td>
<td>50.2%</td>
<td>11.7%</td>
</tr>
<tr>
<td>No. of Groups</td>
<td>21.3</td>
<td>22.2</td>
</tr>
<tr>
<td>Size of Groups</td>
<td>7.7 deg</td>
<td>5.2 deg</td>
</tr>
<tr>
<td>No. of Items</td>
<td>86.1</td>
<td>37.0</td>
</tr>
<tr>
<td>Item Uncertainty</td>
<td>7.2 bits</td>
<td>0.8 bits</td>
</tr>
</tbody>
</table>
have been virtually impossible due to the interrelated nature of the measures. Instead, the displays were designed to keep the correlations among the six measures to a minimum. Table 5 shows the correlations among the six display measures for the 520 displays.

As Table 5 shows, the highest correlation was between overall density and number of items ($r = .97$). This is quite natural and no attempt was made to reduce it. This is analogous to a correlation between number of characters per page and number of words per page in a book. None of the other correlations had an absolute value greater than .73. In general, the correlations make sense. For example, higher local density is associated with larger size groups ($r = .73$) because there are fewer blank spaces on the display to segregate the characters into smaller groups.

4.2.2 Apparatus

The entire study was conducted on an IBM Personal Computer. The displays were shown on an Amdek Model 300 green-phosphor monitor connected to the computer. The monitor displayed 24 lines of 80 columns each. It was adjusted so that the width of the display area was 200 mm and the height was 120 mm. The distance between character centers was 2.5 mm horizontally and 5.0 mm vertically.

4.2.3 Subjects

Ten Bell Laboratories employees participated in the study for about four hours each. Eight of the subjects were clerical employees and the other two were technical and assistant technical. Nine
TABLE 5. Correlations among Six Display Measures for 520 Displays Used in Experiment 1

<table>
<thead>
<tr>
<th></th>
<th>Overall Density</th>
<th>Local Density</th>
<th># of Density Groups</th>
<th># of Density Groups</th>
<th># of Items</th>
<th>Item Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Density</td>
<td>--</td>
<td>.65</td>
<td>-.04</td>
<td>.51</td>
<td>.97</td>
<td>.70</td>
</tr>
<tr>
<td>Local Density</td>
<td>--</td>
<td>--</td>
<td>-.64</td>
<td>.73</td>
<td>.58</td>
<td>.46</td>
</tr>
<tr>
<td>No. of Groups</td>
<td>--</td>
<td>--</td>
<td>-.59</td>
<td>.04</td>
<td>-.05</td>
<td></td>
</tr>
<tr>
<td>Size of Groups</td>
<td>--</td>
<td>--</td>
<td>.45</td>
<td>.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of Items</td>
<td>--</td>
<td>--</td>
<td>.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item Uncertainty</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
of the subjects were female, and nine used corrective lenses. The subjects ranged in age from 24 to 58 years, with a mean of 34. The eight clerical employees had less than one year of experience with the Bell System; the two technical employees had 9 years and 34 years. The eight clerical employees had little previous experience in working with computers, while the two technical employees had occasionally used computers for several years, although neither had programming experience.

4.2.4 Procedure

Each subject saw ten different examples of each format and answered a question about each example. Thus, each subject saw 520 displays and answered as many questions. The same kinds of questions were used for each set of ten displays in one format. The questions used are illustrated in Table 6.

The subjects saw the airline displays in one block and the motel displays in another. Half of the subjects saw the airline displays first and half saw the motel displays first. At the beginning of each block (airline or motel) the subjects were first shown a "practice" format whose data were not used in the analysis. The 26 formats for that block (airline or motel) were then shown in a random order that was determined individually for each subject.

For each format, the subject was first shown an example to study. That was followed by ten examples of the format (each with different data) and a question to answer about each. The question was always shown first. Then, when the subject hit the space bar on the keyboard, the question disappeared and the display appeared.
TABLE 6. Examples of Questions Asked for Airline and Motel Displays in Experiment 1

<table>
<thead>
<tr>
<th>Type of Display</th>
<th>Example of Question</th>
<th>Number Asked per Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airline</td>
<td>What is the coach fare from Baltimore?</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>What is the first class fare from Nashville?</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>What is the arrival time of the 6:50p flight from Cincinnati?</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>What is the flight number of the 3:10p flight from Houston?</td>
<td>3</td>
</tr>
<tr>
<td>Motel</td>
<td>What is the area code for Ashland?</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>What is the phone number (first 3 digits) of the Holiday Inn in Concord?</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>How much is a single room at the Ramada Inn in Charleston?</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>How much is a double room at the Howard Johnsons in Newark?</td>
<td>3</td>
</tr>
</tbody>
</table>
The subject searched the display for the answer to the question and then hit the space bar again when it was found. That caused the display to disappear and the question to reappear, along with two multiple-choice answers. The subject then chose one of the answers. They were encouraged to be as accurate as possible. Search time was measured from appearance of the display until the subject pressed the space bar making the display disappear.

After presentation of the ten examples for one format, the subjects were asked to rate how easy it was to use that format on a scale of 1 to 5, with 1 being "Very Easy to Use" and 5 "Very Difficult to Use". After the subjects had seen all 26 formats in a block (airline or motel) they were given paper copies of one example from each format (in a random order) and asked to sort the pages into five stacks depending upon ease of use. This was simply an additional rating like the one that was done immediately after presentation of that format on the computer, but the paper copies allowed comparisons between formats.

4.3 Results

The results for the two main dependent measures, search time and subjective ratings, will be presented separately.

4.3.1 Search Time

Overall, the subjects answered 97.7% of the questions about the displays correctly, indicating that they followed the instructions to be very accurate. A mean search time was calculated across subjects for each format based on the correct trials only. All subse-
quent analyses were performed on these mean search times. The simple correlations of the six display measures with mean search times are shown in Table 7.

The average size of the groups was the measure that had the highest simple correlation with search time \( (r = .56) \). The scatterplot of these two variables and their linear regression are illustrated in Figure 23. In general, search time increased as the size of the groups on the display increased.

Multiple regressions using various subsets of the six display measures to predict search time were also conducted. Figure 24 plots \( R^2 \) for several multiple regressions as a function of the number of variables in the regression equation. These regressions used a technique known as "regressions by leaps and bounds" (Furnival & Wilson, 1974) to find the best-fitting regression at each number of predictor variables. As Figure 24 shows, the largest increase in \( R^2 \) occurred between one and two predictor variables: the addition of the "number of groups" predictor to the "size of groups" predictor resulted in an increase of \( R^2 \) from .314 to .418. All of the subsequent regressions found that the best fits always included those two predictor variables. Table 8 shows the regression equations for the best fit at each number of predictor variables from two to six.

As Table 8 shows, the multiple regression using all six of the display measures resulted in a multiple \( R^2 \) of .508, meaning that the regression accounted for 50.8% of the variance in search time. Interestingly, the regression using the five display measures besides item uncertainty also resulted in an \( R^2 \) of .508, meaning
### TABLE 7. Experiment 1: Simple Correlations of Six Display Measures with Search Times and Subjective Ratings

<table>
<thead>
<tr>
<th></th>
<th>Search Time</th>
<th>Subjective Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Density</td>
<td>.33</td>
<td>.63</td>
</tr>
<tr>
<td>Local Density</td>
<td>.39</td>
<td>.65</td>
</tr>
<tr>
<td>Number of Groups*</td>
<td>-.06</td>
<td>-.16</td>
</tr>
<tr>
<td>Size of Groups</td>
<td>.56</td>
<td>.67</td>
</tr>
<tr>
<td>Number of Items</td>
<td>.34</td>
<td>.66</td>
</tr>
<tr>
<td>Item Uncertainty</td>
<td>.42</td>
<td>.75</td>
</tr>
</tbody>
</table>

* The correlations with number of groups are low because the functions relating number of groups with search time and subjective rating appear to be non-linear. In both cases, the function is approximately U-shaped, with the optimum search time or rating falling between 20 and 40 groups. However, an analysis of the number of groups in conjunction with the size of those groups indicated that the displays with small numbers of groups tended to have larger sized groups. Thus, the higher search times and ratings for displays with small numbers of groups seem to be due to the size of those groups rather than their number.
Figure 23. Experiment 1: Scatterplot of search time and size of groups
Figure 24. Experiment 1: Plot of $R^2$ for multiple regressions using different numbers of display measures to predict search time
<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search Time =</td>
<td>2.7313</td>
<td>14.91 **</td>
</tr>
<tr>
<td>+ .0107 x Number of Groups</td>
<td>3.03 **</td>
<td></td>
</tr>
<tr>
<td>+ .0896 x Size of Groups</td>
<td>5.90 **</td>
<td></td>
</tr>
<tr>
<td>$\text{R}^2 = .418$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search Time =</td>
<td>2.3171</td>
<td>5.49 **</td>
</tr>
<tr>
<td>+ .0093 x Local Density</td>
<td>1.09</td>
<td></td>
</tr>
<tr>
<td>+ .0123 x Number of Groups</td>
<td>3.23 **</td>
<td></td>
</tr>
<tr>
<td>+ .0784 x Size of Groups</td>
<td>4.28 **</td>
<td></td>
</tr>
<tr>
<td>$\text{R}^2 = .432$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search Time =</td>
<td>1.6696</td>
<td>3.38 **</td>
</tr>
<tr>
<td>- .0209 x Overall Density</td>
<td>-2.28 *</td>
<td></td>
</tr>
<tr>
<td>+ .0286 x Local Density</td>
<td>2.43 *</td>
<td></td>
</tr>
<tr>
<td>+ .0201 x Number of Groups</td>
<td>4.01 **</td>
<td></td>
</tr>
<tr>
<td>+ .0914 x Size of Groups</td>
<td>4.95 **</td>
<td></td>
</tr>
<tr>
<td>$\text{R}^2 = .488$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search Time =</td>
<td>1.7020</td>
<td>3.47 **</td>
</tr>
<tr>
<td>- .0529 x Overall Density</td>
<td>-2.09 *</td>
<td></td>
</tr>
<tr>
<td>+ .0282 x Local Density</td>
<td>2.41 *</td>
<td></td>
</tr>
<tr>
<td>+ .0184 x Number of Groups</td>
<td>3.61 **</td>
<td></td>
</tr>
<tr>
<td>+ .0917 x Size of Groups</td>
<td>5.01 **</td>
<td></td>
</tr>
</tbody>
</table>
Search Time =  
+ .0106 x Number of Items  
\[ \text{R}^2 = .508 \]

\[ \begin{align*}
\text{Search Time} &= 1.5106 \\
- .0515 \times \text{Overall Density} &= -1.97 \\
+ .0284 \times \text{Local Density} &= 2.40 (* \\
+ .0184 \times \text{Number of Groups} &= 3.56 (** \\
+ .0895 \times \text{Size of Groups} &= 4.44 (** \\
+ .0097 \times \text{Number of Items} &= 1.13 \\
+ .0318 \times \text{Item Uncertainty} &= 0.26 \\
\end{align*} \]

\[ \text{R}^2 = .508 \]

** p < .01 (two-tailed) 
* p < .05 (two-tailed)
that item uncertainty did not contribute to the prediction. Individual \( t \) tests on the coefficients from this regression showed that the coefficients for number of groups and size of groups were significant \( (p < .01) \), while the coefficients for overall density and local density were marginally significant \( (p < .05) \), and the coefficient for number of items was not significant. The best-fitting regression using four variables as predictors (overall density, local density, number of groups, size of groups) resulted in a multiple \( R^2 \) of .488, with a similar pattern of significances for the coefficients. The best-fitting regression using three variables as predictors (local density, number of groups, size of groups) resulted in a multiple \( R^2 \) of .432, with only the coefficients for number of groups and size of groups significant. Overall, the results seem to indicate that the two grouping measures are the most important predictors of search time.

In order to choose the best model for predicting search time, a reasonable approach would be to select the model with the smallest number of predictor variables for which no other model has a significantly higher \( R^2 \). That would be the four-variable model shown in Table 8, since the increase in \( R^2 \) from .488 to .508 is not significant, \( t (45) = 1.13 \). The results of the regression using these four predictors are graphically depicted in Figure 25.

4.3.2 Subjective Ratings

A mean subjective rating on the 1 to 5 scale was calculated for each format based on the two ratings that each of the ten subjects provided. Subjective ratings tended to be positively correlated
Figure 25. Experiment 1: Plot of actual search times versus search times predicted from a multiple regression using overall density, local density, number of groups, and size of groups.
with search times ($r = .68$), as one would expect. Higher (worse) subjective ratings were associated with longer search times.

The simple correlations of the six display measures with the mean subjective ratings are shown in Table 7. Most of the correlations were rather high, except for the correlation with number of groups. Interestingly, unlike the search time data, the measure that had the highest correlation with subjective rating was item uncertainty ($r = .75$). The scatterplot of these two variables and their linear regression are illustrated in Figure 26.

The results of a set of regressions using the "leaps and bounds" technique are shown in Figure 27, which plots $R^2$ as a function of the number of predictor variables used. The largest increase in $R^2$ occurred between one and two predictor variables: the addition of the "local density" predictor to the "item uncertainty" predictor resulted in an increase of $R^2$ from $.563$ to $.682$. All of the subsequent regressions found that the best fit always included these two predictors. Table 9 shows the regression equations for the best fit at each number of predictor variables from two to six.

As Table 9 shows, the regression using all six of the display measures resulted in a multiple $R^2$ of $.805$, meaning that these six display measures accounted for 80.5% of the variance in rated ease of use. This degree of fit is quite impressive, and indicates that the display measures are better at predicting subjective ratings than at predicting search times ($R^2$ of $.805$ vs $.508$). In addition, individual $t$ tests on the regression coefficients showed that all of the coefficients were significant (one marginally), unlike the results for search time. This indicates that all six measures play
Figure 26. Experiment 1: Scatterplot of subjective rating and item uncertainty
Figure 27. Experiment 1: Plot of $R^2$ for multiple regressions using different numbers of display measures to predict subjective rating.
<table>
<thead>
<tr>
<th>Coefficient</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating = -3.0593</td>
<td>-4.87 **</td>
</tr>
<tr>
<td>+ .0292 x Local Density</td>
<td>4.27 **</td>
</tr>
<tr>
<td>+ .6102 x Item Uncertainty</td>
<td>6.29 **</td>
</tr>
<tr>
<td>$R^2 = .682$</td>
<td></td>
</tr>
</tbody>
</table>

| Rating = -3.3520                    | -5.34 ** |
| + .0420 x Local Density             | 4.52 ** |
| + .0085 x Number of Groups          | 1.96    |
| + .5374 x Item Uncertainty          | 5.31 ** |
| $R^2 = .706$                        |       |

| Rating = -2.4748                    | -3.76 ** |
| + .0334 x Local Density             | 3.66 ** |
| + .0137 x Number of Groups          | 3.10 ** |
| + .0604 x Size of Groups            | 2.90 ** |
| + .3962 x Item Uncertainty          | 3.73 ** |
| $R^2 = .750$                        |       |

| Rating = -3.8402                    | -4.67 ** |
| - .0258 x Overall Density           | -2.54 * |
| + .0538 x Local Density             | 4.56 ** |
| + .0208 x Number of Groups          | 4.14 ** |
| + .0655 x Size of Groups            | 3.30 ** |
\[ R^2 = .781 \]

\[ \text{Rating} = -3.1627 - .0801 \times \text{Overall Density} + .0522 \times \text{Local Density} + .0184 \times \text{Number of Groups} + .0729 \times \text{Size of Groups} + .0194 \times \text{Number of Items} + .4121 \times \text{Item Uncertainty} \]

\[ R^2 = .805 \]

**p < .01 (two-tailed)**

* p < .05 (two-tailed)
a role in predicting subjective ratings. It appears, however, that
the two most important predictors of subjective rating are local
density and item uncertainty, since those two predictors alone
account for 68.2% of the variance in subjective ratings.

As before, a reasonable approach to choosing the best model for
prediction would be to choose the model with the smallest number of
predictor variables for which no other model has a significantly
higher \( R^2 \). That would be the complete six-variable model shown in
Table 9, since the increase in \( R^2 \) from the five-variable model
(.781) to the six-variable model (.805) is significant,
\( t(45) = 2.35, p < .05 \). The results of this regression using all
six variables are graphically depicted in Figure 28.

4.4 Discussion

Overall, the results of Experiment 1 were quite encouraging.
The fact that the regressions accounted for 51% of the variance in
search times and 80% of the variance in subjective ratings indi-
cates that it may be possible to predict the usability of a display
based on these display measures.

Several interesting results from Experiment 1 warrant further
discussion:

1. Clearly, the subjective ratings of the displays were better
   predicted by the display measures than were the search times
   \( (R^2 \) of .805 vs .508). This is not really very surprising
   when one considers the source of the display measures used in
   the regressions. Those measures were designed to assess
characteristics of displays that published guidelines (and,
Figure 28. Experiment 1: Plot of actual subjective ratings versus subjective ratings predicted from a multiple regression using overall density, local density, number of groups, size of groups, number of items, and item uncertainty.
to a lesser extent, empirical data) claimed are important. For the most part, these guidelines are simply their authors' ideas about what makes a "good" display. The fact that 80% of the variance in subjective ratings could be accounted for using these measures means two things, then: (1) The display measures as implemented in the "C" program successfully captured the characteristics of displays that the authors of the guidelines said are important; and (2) Those authors' subjective reactions to displays are similar to those of the subjects in this experiment.

2. The multiple regressions indicate that different display variables are important in predicting search times vs subjective ratings. This is apparent from a comparison of Figures 24 and 27. The single best predictor of search times was the average size of the groups on the display ($r^2 = .314$); the single best predictor of subjective ratings was the average uncertainty of an item's location ($r^2 = .563$). Interestingly, this item uncertainty measure was of little importance in predicting search time, since adding it to the regression containing the other five variables did not change $R^2$. The two best predictors of search times were number of groups and size of groups ($R^2 = .418$); the two best predictors of subjective rating were local density and item uncertainty ($R^2 = .682$). In short, search times are most closely related to the way characters are grouped on the display, while subjective ratings are most closely related to how "tightly packed" the display is (local density) and whether the items
are aligned vertically and horizontally (item uncertainty).

3. The best models for predicting search times and subjective ratings resulted in negative coefficients for overall density. (See Tables 8 and 9.) This is rather surprising since it means that if the other display variables were held constant, an increase in overall density would result in a lower (improved) predicted search time and subjective rating. This is contrary to all the previous visual search studies which found that increasing overall density degrades performance. However, it must be kept in mind that the present experiment did find a positive simple correlation of overall density with both search time and subjective rating (Table 7). The reason for such a suppressor relationship arising in the multiple regression is not clear. Speculation as to the origin of such a relationship is difficult because it is hard to imagine two displays that have different overall densities but similar values for the other measures.
5. EXPERIMENT 2: VALIDATION OF THE PREDICTION SYSTEM

5.1 Introduction

The results of Experiment 1 indicated that it may be possible to predict search times and subjective ratings of displays based on regression equations that use certain display measures. The major question that remains is how well these regression equations will generalize to other displays and other subjects. Experiment 2 addressed that question by using the regression equations developed in Experiment 1 (Tables 8 and 9) to predict, a priori, the search times and subjective ratings for a completely new set of displays and new group of subjects.

5.2 Method

5.2.1 Displays

Fifteen different formats for presenting listings of data about books were developed. On each display, authors were listed alphabetically by last name, and, for each author, books were listed alphabetically by title. Additional information included an arbitrary author number, the price of the book, publisher, and number of pages. Two examples of the book displays are shown in Figure 29.

In order to reduce any tendency to make these book displays similar in format to the airline and motel displays (and thus bias the results), over half of the book formats (eight) were designed by people who were not familiar with the displays used in Experiment 1. These eight formats were designed by four acquaintances of
<table>
<thead>
<tr>
<th>Books</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author:</strong></td>
</tr>
<tr>
<td><strong>Author#:</strong></td>
</tr>
<tr>
<td><strong>Title:</strong></td>
</tr>
<tr>
<td><strong>Price:</strong></td>
</tr>
<tr>
<td><strong>Publisher:</strong></td>
</tr>
<tr>
<td><strong>Pages:</strong></td>
</tr>
<tr>
<td><strong>Author:</strong></td>
</tr>
<tr>
<td><strong>Author#:</strong></td>
</tr>
<tr>
<td><strong>Title:</strong></td>
</tr>
<tr>
<td><strong>Price:</strong></td>
</tr>
<tr>
<td><strong>Publisher:</strong></td>
</tr>
<tr>
<td><strong>Pages:</strong></td>
</tr>
<tr>
<td><strong>Author:</strong></td>
</tr>
<tr>
<td><strong>Author#:</strong></td>
</tr>
<tr>
<td><strong>Title:</strong></td>
</tr>
<tr>
<td><strong>Price:</strong></td>
</tr>
<tr>
<td><strong>Publisher:</strong></td>
</tr>
<tr>
<td><strong>Pages:</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Books</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Silverberg, R</strong></td>
</tr>
<tr>
<td><strong>Silverberg, R</strong></td>
</tr>
<tr>
<td><strong>Silverberg, R</strong></td>
</tr>
<tr>
<td><strong>Silverberg, R</strong></td>
</tr>
<tr>
<td><strong>Silverberg, R</strong></td>
</tr>
<tr>
<td><strong>Silverberg, R</strong></td>
</tr>
<tr>
<td><strong>Springer, N</strong></td>
</tr>
<tr>
<td><strong>Springer, N</strong></td>
</tr>
<tr>
<td><strong>Springer, N</strong></td>
</tr>
<tr>
<td><strong>Stewart, M</strong></td>
</tr>
<tr>
<td><strong>Stewart, M</strong></td>
</tr>
</tbody>
</table>

**Figure 29.** Two examples of book displays used in Experiment 2.
the author who had no prior experience in computer system design or human interface design. Each of them was asked to design (on coding forms) two ways of presenting the specified book data. The only other information provided was the maximum length of each of the data items (e.g., author, title, etc.). The original intention was to have all of the displays in Experiment 2 designed by people other than the author, but the display formats designed by other people tended to be similar to each other. Consequently, the author designed the remaining seven formats in order to introduce more variation and provide a better test of the prediction system.

For each of the fifteen display formats, ten examples containing different data were developed. Thus, a total of 150 displays were used in the experiment. To illustrate the resulting values for the six display measures, Table 10 shows their means, standard deviations, and ranges. A comparison of this table to Table 4, which gives similar data for Experiment 1, shows that, overall, the displays used in the two experiments did not differ drastically on most of the measures. The largest difference was in mean number of items (86.1 in Experiment 1 vs. 58.6 in Experiment 2).

For each of the fifteen display formats, search times and subjective ratings were predicted using the best models from Experiment 1. For search times, that was the four-predictor model shown in Table 8. For subjective ratings, that was the six-predictor model shown in Table 9. The mean of the predicted search times was 3.69 sec and the range was 3.03 sec to 4.64 sec. The mean of the predicted subjective ratings was 2.46 and the range was 0.82 to 4.78. The minimum value of that range for the subjective ratings
TABLE 10. Means, Standard Deviations, and Ranges of Six Display Measures for 150 Displays Used in Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Density</td>
<td>20.3%</td>
<td>11.9%</td>
<td>4.5% - 56.7%</td>
</tr>
<tr>
<td>Local Density</td>
<td>46.8%</td>
<td>11.7%</td>
<td>27.5% - 67.5%</td>
</tr>
<tr>
<td>Number of Groups</td>
<td>19.1</td>
<td>19.5</td>
<td>1 - 67</td>
</tr>
<tr>
<td>Size of Groups</td>
<td>7.9 deg</td>
<td>5.3 deg</td>
<td>2.3 deg - 25.5 deg</td>
</tr>
<tr>
<td>Number of Items</td>
<td>58.6</td>
<td>36.4</td>
<td>13 - 169</td>
</tr>
<tr>
<td>Item Uncertainty</td>
<td>6.7 bits</td>
<td>1.4 bits</td>
<td>4.2 bits - 9.3 bits</td>
</tr>
</tbody>
</table>
(0.82) points out an interesting aspect of the predictions: three of the display formats had predicted ratings under one. The rating scale used by the subjects, however, had a possible range of only one to five. Thus, the predictions under one will by definition be wrong. The fact that the predicted ratings can be less than one, and even less than zero, is apparent from Table 9, which shows that the regression equations have negative intercept values. This points out a basic difference between the search times and subjective ratings—a difference that reflects a possible flaw in the subjective ratings. The subjective ratings are really relative comparisons of the displays in a set to each other; thus, the rating given to one display depends upon the characteristics of the other displays in the set. The search times, on the other hand, may be considered more "stable" measures since they are not dependent on the characteristics of the entire set of displays.

5.2.2 Apparatus

The apparatus was the same as in Experiment 1, and the monitor was adjusted to the same display parameters.

5.2.3 Subjects

Fourteen Bell Laboratories employees participated in the study for about one hour each. All of the subjects were either assistant technical or senior assistant technical. Five of the subjects were female, and eight used corrective lenses. The subjects ranged in age from 23 to 45 years, with a mean of 30. Their experience with the Bell System ranged from 2 to 23 years, with a mean of 7.3. All
of the subjects had prior experience in working with computers, most for several years. Eight of the subjects had programming experience.

In general, the subjects in Experiment 2 were much more experienced in working with computers than were those in Experiment 1. This was not really intentional, but rather a function of what subjects were available when the two experiments were run. This difference provided a test of how well the regression equations from Experiment 1 would generalize to a different subject population.

5.2.4 Procedure

Each subject saw ten different examples of each display format and answered a question about each example. Thus, each subject saw 150 displays and answered as many questions. The same basic set of questions was used for each set of ten displays in one format. The questions used are illustrated in Table 11.

The rest of the procedure was basically the same as in Experiment 1 except that only fifteen different formats were used and they were presented in one block. As in Experiment 1, search times were recorded on each trial. Subjective ratings were obtained at the end of each format and again at the end of the experiment.
TABLE 11. Examples of Questions Asked for Book Displays in Experiment 2

<table>
<thead>
<tr>
<th>Example of Question</th>
<th>Number Asked per Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>How many pages is Under the Lilacs by Alcott, L M?</td>
<td>2</td>
</tr>
<tr>
<td>What is the author number for Silverberg, R?</td>
<td>3</td>
</tr>
<tr>
<td>Who is the publisher of Murderess Ink by Winn, D?</td>
<td>2</td>
</tr>
<tr>
<td>What is the price of Genesis Machine by Hogan, J P?</td>
<td>3</td>
</tr>
</tbody>
</table>
5.3 Results

5.3.1 Search Times

Overall, the subjects answered 98.8% of the questions about the displays correctly, reflecting a high level of accuracy similar to that in Experiment 1. A mean search time was calculated across subjects for each format based on the correct trials only. The result of most interest is the simple correlation between these actual mean search times and the search times predicted using the four-predictor regression equation from Experiment 1 (Table 8). The four-variable model predicted search times quite well: \( r = .800 \). The scatterplot of actual search times and search times predicted using this model is shown in Figure 30.

Figure 30 also shows the regression line that provides the best fit between predicted and actual search times. If this were a perfect prediction system, the intercept of that line would be 0 and the slope would be 1. The actual intercept of -0.62 and slope of 1.06 are surprisingly close to those "perfect" values. As the standard errors of the intercept (0.82) and slope (.22) reflect, the actual values do not significantly differ from 0 and 1, respectively.

The most surprising finding was the high value of \( r^2 \) (.64). The original regression from Experiment 1 that used overall density, local density, number of groups, and size of groups had an \( R^2 \) of .49. (See Table 8.) One would expect, in a validation study such as Experiment 2, that \( r^2 \) would be no greater than the corresponding \( R^2 \) found in the original experiment. However, apparently all the
Figure 30. Experiment 2: Scatterplot of actual search times and search times predicted from a model developed in Experiment 1 that uses overall density, local density, number of groups, and size of groups. Letters indicate different displays.
displays used in Experiment 2 fit into the original model quite well, resulting in a higher $r^2$ (i.e., there were not any significant outliers).

As additional confirmation of the predictive validity of the various display measures, Table 12 shows the simple correlations between the actual search times and the search times predicted by all of the models listed in Table 8. As expected from Experiment 1, the predictions from the four-variable model had the highest correlation with actual search times. More important, however, is that all of the correlations were quite high (.713 to .800).

5.3.2 Subjective Ratings

A mean subjective rating on the 1 to 5 scale (1 = Very Easy to Use, 5 = Very Difficult to Use) was calculated for each format based on the two ratings that each of the fourteen subjects provided. As in Experiment 1, the subjective ratings were positively correlated with search times ($r = .80$).

The result of most interest is the simple correlation between actual subjective ratings and the subjective ratings predicted using the six-variable model from Experiment 1 (Table 9). This correlation was quite high: $r = .799$. The scatterplot of actual subjective ratings and subjective ratings predicted using this model is shown in Figure 31.

Figure 31 also shows the regression line that provides the best fit between predicted and actual subjective ratings. As with the search times, perfect prediction would yield an intercept of 0 and slope of 1. The actual intercept of 1.32 and slope of 0.40 diverge
TABLE 12. Simple Correlations Between Actual Search Times from Experiment 2 and Predicted Search Times Based on Five Different Regression Equations from Experiment 1

<table>
<thead>
<tr>
<th>Display Variables Used in Regression</th>
<th>Correlation (r) of Actual and Predicted Search Times</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Groups, Size of Groups</td>
<td>.713</td>
</tr>
<tr>
<td>Local Density, Number of Groups,</td>
<td></td>
</tr>
<tr>
<td>Size of Groups</td>
<td>.759</td>
</tr>
<tr>
<td>Overall Density, Local Density,</td>
<td></td>
</tr>
<tr>
<td>Number of Groups, Size of Groups</td>
<td>.800</td>
</tr>
<tr>
<td>Overall Density, Local Density,</td>
<td></td>
</tr>
<tr>
<td>Number of Groups, Size of Groups,</td>
<td></td>
</tr>
<tr>
<td>Number of Items</td>
<td>.795</td>
</tr>
<tr>
<td>Overall Density, Local Density,</td>
<td></td>
</tr>
<tr>
<td>Number of Groups, Size of Groups,</td>
<td></td>
</tr>
<tr>
<td>Number of Items, Item Uncertainty</td>
<td>.796</td>
</tr>
</tbody>
</table>
Figure 31. Experiment 2: Scatterplot of actual subjective ratings and subjective ratings predicted from a regression equation developed in Experiment 1 that uses overall density, local density, number of groups, size of groups, number of items, and item uncertainty. Letters indicate different displays.
from those perfect values. As the standard errors of the intercept (0.23) and slope (0.08) indicate, the values differ significantly from 0 and 1, respectively.

The reason for the difference in slope and intercept values does not lie primarily in a difference between the means of the ratings (2.46 for predicted vs. 2.31 for actual) but rather a difference between the ranges. The range for the actual subjective ratings (1.36 to 3.46) is more restricted than the range for the predicted subjective ratings (0.82 to 4.78). It is difficult to attribute this difference to some characteristics of the displays themselves, since the subjective rating scale should be "self-adjusting" (i.e., the score given to any one display format should be determined by how that format compares to the other formats in the experiment). It seems more likely that the difference in range is due to the different population of subjects represented in Experiment 2 compared to Experiment 1 (i.e., more "computer-sophisticated" users). Perhaps their wider experience with computer systems and CRT displays, both good and bad, led them to view these displays as covering a relatively small part of the "good-to-bad" continuum.

Unlike the results for search time, $R^2$ for subjective ratings decreased in Experiment 2 (.638) relative to the corresponding $R^2$ in Experiment 1 (.805). This shrinkage in $R^2$ is larger than would have been expected based upon Experiment 1. The so-called "shrunk $R^2" can be derived from Experiment 1 to provide an unbiased estimate of the population $R^2$. (See, for example, Cohen & Cohen, 1975, p. 106.) That shrunk $R^2$ estimated from Experiment 1 is .779, which is still larger than $R^2$ from Experiment 2. This is,
perhaps, additional evidence for the effect of different subject populations in Experiments 1 and 2.

Further confirmation of the predictive validity of the display measures is provided by Table 13, which shows the simple correlations with subjective rating for all of the models listed in Table 9. As expected from Experiment 1, the predictions from the six-variable model had the highest correlation with actual subjective ratings. However, as with the search times, all of the correlations were quite high (.723 to .799).

5.4 Discussion

The results of Experiment 2 showed that the prediction system developed in Experiment 1 generalizes quite well to a different set of displays and different subject population. In particular, the high correlation between predicted and actual search times \( r = .800 \) was most encouraging. Likewise, the high correlation between predicted and actual subjective ratings \( r = .799 \) was also encouraging, but it was less than expected based upon the results of Experiment 1.

Another way of looking at the accuracy of the prediction system is to consider how it would be used in a real application. Typically, a system designer's task is to choose which of several alternative display formats is best. What is meant by "best" depends on the circumstances. If the system is to be used in a situation where speed is important (e.g., telephone directory assistance) then search time might be emphasized. If the system is to be marketed to the general public (e.g., home information
<table>
<thead>
<tr>
<th>Display Variables Used in Regression</th>
<th>Correlation ($r$) of Actual and Predicted Subjective Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local Density, Item Uncertainty</td>
<td>.734</td>
</tr>
<tr>
<td>Local Density, Number of Groups,</td>
<td>.723</td>
</tr>
<tr>
<td>Item Uncertainty</td>
<td></td>
</tr>
<tr>
<td>Local Density, Number of Groups,</td>
<td>.756</td>
</tr>
<tr>
<td>Size of Groups, Item Uncertainty</td>
<td></td>
</tr>
<tr>
<td>Overall Density, Local Density,</td>
<td>.772</td>
</tr>
<tr>
<td>Number of Groups, Size of Groups,</td>
<td></td>
</tr>
<tr>
<td>Item Uncertainty</td>
<td></td>
</tr>
<tr>
<td>Overall Density, Local Density,</td>
<td>.799</td>
</tr>
<tr>
<td>Number of Groups, Size of Groups,</td>
<td></td>
</tr>
<tr>
<td>Number of Items, Item Uncertainty</td>
<td></td>
</tr>
</tbody>
</table>
retrieval) then subjective ratings might be emphasized. In many cases, however, the system designer would probably want to give equal weights to search time and subjective rating. One way of doing that would be to convert the search times and subjective ratings to Z scores and then simply add them together. A scatterplot of the predicted and actual scores transformed in this manner is shown in Figure 32.

Since a system designer would rarely have empirical data on the usability of alternative formats, he or she could only rely on predicted scores to guide the display selection. Based on Figure 32, the system designer faced with choosing the best format for these book displays would choose format "i" since it yielded the lowest predicted score (-2.66). (This is also relatively obvious from Figures 30 and 31.) Inspection of the actual scores shows that format "i" would have been the best choice, since it yielded the lowest actual score as well.

Another reasonable choice based on the predicted scores would have been format "j", since its score (-2.64) was quite close to the lowest one. Likewise, it would have been a reasonably good choice since its actual score was one of the three lowest. In short, the prediction system seems to perform well when used in the manner that it might be in a real system design application.
Figure 32. Experiment 2: Scatterplot of actual transformed scores and predicted transformed scores based on regression equations developed in Experiment 1. Transformed scores are the sum of z-score transformations for search times and subjective ratings. Letters indicate different displays.
6. GENERAL DISCUSSION: APPLYING THE PREDICTION SYSTEM TO DISPLAYS IN THE LITERATURE

While the results of Experiments 1 and 2 were quite impressive, one might still wonder how well this prediction system would generalize to other types of displays and tasks. In particular, the task in both Experiments 1 and 2 involved searching an ordered list of items for a particular piece of information. In the set of real-world displays, however, this is often not the case. It would be informative to apply the prediction system developed using these displays of ordered lists to other types of displays reported in the literature.

As described earlier (Section 2.2) there is a relatively small set of empirical studies of realistic displays. Of those studies listed earlier in Table 2, only four yielded data that can be directly compared to search times from the prediction system (i.e., Tullis, 1981; Dodson & Shields, 1978; Ringel & Hammer, 1964; Callan et al., 1977). The remaining studies of realistic displays listed in Table 2 either did not manipulate display format (e.g., Card, 1982; Vartabedian, 1971) or did not use a search task (e.g., Grace, 1966; Wright, 1977).

An obvious starting point in the comparison of predictions to data in the literature is the Tullis (1981) study, which provided the impetus for this entire line of research. As described earlier (Section 1.2), and as shown in Table 14, that study found that the "narrative" format (Figure 1) required 8.3 sec to interpret, while the "structured" format (Figure 2) required 5.0 sec. Using the four-predictor model from Table 8 results in predicted search times
### TABLE 14. Results of Applying the Prediction System to Search Time Data in the Literature

<table>
<thead>
<tr>
<th>Reference</th>
<th>Display Conditions</th>
<th>Actual</th>
<th>Real</th>
<th>Adj*</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tullis (1981)</td>
<td>Narrative</td>
<td>8.3</td>
<td>4.2</td>
<td>-</td>
<td>Same order</td>
</tr>
<tr>
<td></td>
<td>Structured</td>
<td>5.0</td>
<td>3.2</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Dodson &amp; Shields (1978)</td>
<td>&quot;Density&quot;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>30%</td>
<td>3.4</td>
<td>3.0</td>
<td>3.5</td>
<td>$r = .985$</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>3.9</td>
<td>3.1</td>
<td>3.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>70%</td>
<td>5.0</td>
<td>3.9</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>Ringel &amp; Hammer (1964)</td>
<td>&quot;#lines&quot; &quot;density&quot;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 low</td>
<td>18.7</td>
<td>4.8</td>
<td>17.1</td>
<td>$r = .792$</td>
</tr>
<tr>
<td></td>
<td>18 low</td>
<td>17.7</td>
<td>5.9</td>
<td>19.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25 low</td>
<td>22.3</td>
<td>6.8</td>
<td>21.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 med.</td>
<td>16.9</td>
<td>4.8</td>
<td>17.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18 med.</td>
<td>21.1</td>
<td>5.9</td>
<td>19.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25 med.</td>
<td>21.4</td>
<td>6.8</td>
<td>21.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 high</td>
<td>16.2</td>
<td>4.8</td>
<td>17.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18 high</td>
<td>17.7</td>
<td>5.9</td>
<td>19.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25 high</td>
<td>20.1</td>
<td>6.8</td>
<td>21.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 higher**</td>
<td>-</td>
<td>3.5</td>
<td>14.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>higher**</td>
<td>_</td>
<td>3.6</td>
<td>14.7</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>higher**</td>
<td>_</td>
<td>3.6</td>
<td>14.8</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Callan, &quot;#items&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curran, 6</td>
</tr>
<tr>
<td>4.2  2.9  4.6  ( r = .906 )</td>
</tr>
<tr>
<td>&amp; Lane (1977) 10</td>
</tr>
<tr>
<td>4.5  2.8  4.5</td>
</tr>
<tr>
<td>19</td>
</tr>
<tr>
<td>5.0  3.2  4.8</td>
</tr>
<tr>
<td>24</td>
</tr>
<tr>
<td>5.2  3.3  5.0</td>
</tr>
<tr>
<td>33</td>
</tr>
<tr>
<td>5.2  3.6  5.3</td>
</tr>
<tr>
<td>40</td>
</tr>
<tr>
<td>5.5  3.8  5.6</td>
</tr>
</tbody>
</table>

* The "adjusted" predicted search times were derived from a regression in which the actual search times were fit using the "real" predicted search times from the model. The correlation associated with that regression is shown in the last column of the table. The purpose of the adjustment is to take into account the differences between the types of search tasks used in the study being considered and Experiment 1 of the present studies.

** Not included in Ringel and Hammer study or in calculation of correlation. (See text.)
of 4.23 sec for the "narrative" format and 3.20 sec for the "structured" format. Obviously, these predicted times are significantly lower than the actual times found by Tullis. One reason for the difference is that the times measured by Tullis (1981) included the time to select the correct answer from a multiple-choice list, while such time was not included in the experiment that developed the prediction system. Another reason is that some of the times in the earlier study were for questions that involved searching for and encoding multiple items on the display. Thus, one would expect the simple search times predicted by the regression equation to be shorter. The more important finding is that the order of the "narrative" and "structured" formats is the same for the predicted and actual times: both show that the "structured" format is better. In addition, the ratio of "structured" to "narrative" is similar for the predicted (.76) and actual (.60) times.

The displays studied by Tullis (1981) also point out what may be a shortcoming of the prediction system. The grouping algorithm (described in Section 3.4) detected 13 groups of characters in the "structured" format. One of those groups was the box of asterisks at the top of the display. (The two items inside the box were detected as two more groups.) One could argue that the box of asterisks should be ignored in the analysis since it simply acts as a delimiter or highlighting mechanism and does not convey information itself. If the box of asterisks is removed and the predicted search time recomputed, the result is 2.96 sec. This results in a greater difference between the predicted times for the structured and narrative formats, and a ratio (.70) that is closer to the
actual ratio (.60). Although this seems to indicate that "pseudographics" like the box of asterisks should be treated in a special way by the display analysis program, further research is needed to define the most appropriate treatment.

Applying the prediction system to the two formats studied by Tullis (1981) obviously is not a very convincing test since there was an even chance of predicting the correct order for only two formats. A bit more rigorous test is provided by the three display formats studied by Dodson and Shields (1978). They manipulated the overall density of displays for a Space Shuttle application using three levels: 30%, 50%, and 70%. Examples of the three formats are shown in Figures 33 through 35. The display parameters as measured by the analysis program are shown in Table 15 for these three formats. Oddly, Dodson and Shields must have used a different method for measuring overall density, since what they call 30%, 50%, and 70% actually are 24.9%, 41.6%, and 58.8%, respectively, when the total number of characters displayed is expressed as a percentage of the total number of character spaces available. (Dodson and Shields specified that their terminal had 17 user-available lines and 47 character spaces per line.) A more important point about the display formats is that they obviously differed on other parameters besides overall density. For example, the number of groups varied from 4 to 10, while the average size of those groups varied from 16.8 deg to 4.9 deg.

Using the four-variable model from Table 8 to predict search times for these formats results in times of 2.95 sec, 3.09 sec, and 3.88 sec for the so-called 30%, 50%, and 70% formats. These
**Figure 33.** 30% density display studied by Dodson and Shields (1978)

**Figure 34.** 50% density display studied by Dodson and Shields (1978)
<table>
<thead>
<tr>
<th>AEPI COMMANDS</th>
<th>70 PERCENT DSPLY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 EMERGENCY PARK=YES/NO</td>
<td>6 START=YES/NO</td>
</tr>
<tr>
<td>2 MOUNT CHECK=YES/NO</td>
<td>7 EDIT=YES/NO</td>
</tr>
<tr>
<td>3 SW CHECK=YES/NO</td>
<td>8 SELF DUMP=YES/NO</td>
</tr>
<tr>
<td>4 RUN=YES/NO</td>
<td>9 SW RELOAD=YES/NO</td>
</tr>
<tr>
<td>5 HALT=YES/NO</td>
<td>10 CALIBRATE=YES/NO</td>
</tr>
</tbody>
</table>

FO SEQUENCE

| 11 1A-774:12:36 | 15 2C-445:35:43 |
| 12 1B-785:48:50 | 16 03-451:19:42 |
| 14 2B-300:56:41 | 18 05-515:21:20 |

08-288:28:26 432:48:56 DELAYED

| 19 FWO=73      | 23 FOCUS=15  |
| 20 FW1=83      | 24 CAM HV=2  |
| 21 FW2=61      | 25 PCA HV=9  |
| 22 FOV=14      | 26 EXP TIME=1 |

| 27 TRACK=YES/NO| 28 RAY=90    |
| 29 MPD DELAY 88| 30 RESET=703 |

**Figure 35.** 70% density display studied by Dodson and Shields (1978)
<table>
<thead>
<tr>
<th>Display</th>
<th>Overall Density</th>
<th>Local Density</th>
<th># of Groups</th>
<th>Group Size</th>
<th># of Items</th>
<th>Item Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 33</td>
<td>24.9%</td>
<td>40.3%</td>
<td>10</td>
<td>4.9 deg</td>
<td>22</td>
<td>6.91 bits</td>
</tr>
<tr>
<td>Fig. 34</td>
<td>41.6%</td>
<td>55.2%</td>
<td>8</td>
<td>6.0 deg</td>
<td>29</td>
<td>6.46 bits</td>
</tr>
<tr>
<td>Fig. 35</td>
<td>58.8%</td>
<td>63.8%</td>
<td>4</td>
<td>16.8 deg</td>
<td>39</td>
<td>7.38 bits</td>
</tr>
</tbody>
</table>
predicted times, along with the actual times found in the study are listed in Table 14. As with the Tullis (1981) displays, the predicted times are all shorter than their corresponding actual times (although not by much). This is not surprising considering the greater complexity of the Dodson and Shields displays relative to the motel and airline displays used in development of the prediction system. The important result is the high correlation between the values \( r = .985 \).

The regression used to obtain that correlation coefficient can also be used to adjust the predicted times to take into account the greater complexity of the search task used by Dodson and Shields. That adjustment simply involves inserting the predicted times into the regression equation that fits the actual times using the predicted times. These adjusted search times are shown in Table 14. Both the actual search times and the predicted search times, adjusted in this manner, are plotted in Figure 36 by display density, as in the original study.

Dodson and Shields (1978) took their data as being evidence for an exponential function: "As can be seen from this figure, response time [search time] as a function of display density is an exponential function with times rapidly increasing as display density exceeds 60%" (Dodson & Shields, 1978, p. 2-7). This assertion has very little foundation, since it is based on only three data points, and since those three points can be fit quite well by a linear function of display density \( r = .978 \). Their finding, however, apparently led NASA to adopt the following guideline for their Spacelab displays: "Display character density should not
Figure 36. Plot of actual and predicted search times from Dodson and Shields (1978) as a function of display density.
exceed 60% of the available character spaces" (NASA, 1979, p. x). The most interesting point about Figure 36 is that the predicted search times exhibit almost exactly the same "exponential" shape as a function of display density. The critical point to remember is that these predicted times were derived from a linear combination of overall density, local density, number of groups, and size of groups. There is no need to postulate an exponential function.

The prediction system can also be applied to the displays of military information studied by Ringel and Hammer (1964). These displays are somewhat different from those discussed so far in that they were not CRT displays. Ringel and Hammer used typed tables of data describing the status of military units. These were then photographed and shown by a slide projector. An example of such a display is shown in Figure 37. They used nine different display formats determined by the factorial combination of three levels of "amount of information" (10, 18, or 25 lines) and three levels of "density" (low, medium, or high). "Density" was defined as the ratio of the height of a letter to the space between lines: low = 1:4, medium = 1:3, and high = 1:2. Thus, their measure of "amount of information" correlates with overall density and their measure of "density" correlates with local density. The tasks performed with these displays were more complex than simple search, involving extraction of information from multiple columns (e.g., one question might be, "Which unit has an armor equipment status of 95 and is combat experienced?").

The actual search times found in the study and those predicted by the regression equation are shown in Table 14 for all nine con-
<table>
<thead>
<tr>
<th>UNIT</th>
<th>LOCATION</th>
<th>DATE</th>
<th>COMMD</th>
<th>EFFECT PT</th>
<th>STRGTH</th>
<th>ART</th>
<th>TRN</th>
<th>EFF</th>
<th>CMB</th>
<th>PRESENT ACTIVITY</th>
<th>REMARKS</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 INF DIV</td>
<td>XL 21</td>
<td>AUG 62</td>
<td></td>
<td>1400</td>
<td>80</td>
<td>75</td>
<td>95</td>
<td></td>
<td></td>
<td>DEFENDING</td>
<td>FIRST ACTION</td>
</tr>
<tr>
<td>29 INF DIV</td>
<td>TN 11</td>
<td>OCT 62</td>
<td></td>
<td>1300</td>
<td>85</td>
<td>90</td>
<td>95</td>
<td></td>
<td></td>
<td>SCREENING</td>
<td>COMBAT EXPER</td>
</tr>
<tr>
<td>26 INF DIV</td>
<td>LD 19</td>
<td>AUG 62</td>
<td></td>
<td>1400</td>
<td>75</td>
<td>85</td>
<td>90</td>
<td></td>
<td></td>
<td>SCREENING</td>
<td>HIGH MORALE</td>
</tr>
<tr>
<td>31 INF DIV</td>
<td>KR 14</td>
<td>SEPT 62</td>
<td></td>
<td>1500</td>
<td>90</td>
<td>75</td>
<td>85</td>
<td></td>
<td></td>
<td>ATTACKING</td>
<td>MOD RESISTENCE</td>
</tr>
<tr>
<td>28 INF DIV</td>
<td>TN 11</td>
<td>JULY 62</td>
<td></td>
<td>1100</td>
<td>90</td>
<td>90</td>
<td>75</td>
<td></td>
<td></td>
<td>ATTACKING</td>
<td>SWAMPY TERRN</td>
</tr>
<tr>
<td>33 INF DIV</td>
<td>PT 30</td>
<td>OCT 62</td>
<td></td>
<td>1500</td>
<td>85</td>
<td>85</td>
<td>85</td>
<td></td>
<td></td>
<td>ADVANCING</td>
<td>FIRST ACTION</td>
</tr>
<tr>
<td>32 INF DIV</td>
<td>PT 30</td>
<td>NOV 62</td>
<td></td>
<td>1200</td>
<td>85</td>
<td>95</td>
<td>95</td>
<td></td>
<td></td>
<td>ASSEMBLING</td>
<td>MOD RESISTENCE</td>
</tr>
<tr>
<td>36 INF DIV</td>
<td>XL 21</td>
<td>JULY 62</td>
<td></td>
<td>1200</td>
<td>90</td>
<td>80</td>
<td>75</td>
<td></td>
<td></td>
<td>ATTACKING</td>
<td>SWAMPY TERRN</td>
</tr>
<tr>
<td>17 INF DIV</td>
<td>XL 21</td>
<td>OCT 62</td>
<td></td>
<td>1400</td>
<td>90</td>
<td>95</td>
<td>85</td>
<td></td>
<td></td>
<td>ATTACKING</td>
<td>FIRST ACTION</td>
</tr>
<tr>
<td>35 INF DIV</td>
<td>LD 19</td>
<td>OCT 62</td>
<td></td>
<td>1000</td>
<td>80</td>
<td>80</td>
<td>90</td>
<td></td>
<td></td>
<td>ASSEMBLING</td>
<td>HIGH MORALE</td>
</tr>
<tr>
<td>23 INF DIV</td>
<td>XL 21</td>
<td>SEPT 62</td>
<td></td>
<td>1300</td>
<td>80</td>
<td>90</td>
<td>95</td>
<td></td>
<td></td>
<td>ADVANCING</td>
<td>SWAMPY TERRN</td>
</tr>
<tr>
<td>18 INF DIV</td>
<td>KR 14</td>
<td>NOV 62</td>
<td></td>
<td>1500</td>
<td>85</td>
<td>90</td>
<td>75</td>
<td></td>
<td></td>
<td>ASSEMBLING</td>
<td>HIGH MORALE</td>
</tr>
<tr>
<td>30 INF DIV</td>
<td>PT 30</td>
<td>AUG 62</td>
<td></td>
<td>1200</td>
<td>75</td>
<td>85</td>
<td>90</td>
<td></td>
<td></td>
<td>ADVANCING</td>
<td>COMBAT EXPER</td>
</tr>
<tr>
<td>16 INF DIV</td>
<td>PT 30</td>
<td>SEPT 62</td>
<td></td>
<td>1100</td>
<td>75</td>
<td>85</td>
<td>85</td>
<td></td>
<td></td>
<td>ASSEMBLING</td>
<td>SWAMPY TERRN</td>
</tr>
<tr>
<td>27 INF DIV</td>
<td>TN 11</td>
<td>JULY 62</td>
<td></td>
<td>1400</td>
<td>80</td>
<td>75</td>
<td>75</td>
<td></td>
<td></td>
<td>DEFENDING</td>
<td>COMBAT EXPER</td>
</tr>
<tr>
<td>34 INF DIV</td>
<td>KR 14</td>
<td>JULY 62</td>
<td></td>
<td>1100</td>
<td>90</td>
<td>85</td>
<td>75</td>
<td></td>
<td></td>
<td>DEFENDING</td>
<td>FIRST ACTION</td>
</tr>
<tr>
<td>45 INF DIV</td>
<td>KR 14</td>
<td>SEPT 62</td>
<td></td>
<td>1100</td>
<td>95</td>
<td>95</td>
<td>80</td>
<td></td>
<td></td>
<td>ASSEMBLING</td>
<td>COMBAT EXPER</td>
</tr>
<tr>
<td>43 INF DIV</td>
<td>TN 11</td>
<td>SEPT 62</td>
<td></td>
<td>1300</td>
<td>95</td>
<td>90</td>
<td>80</td>
<td></td>
<td></td>
<td>DEFENDING</td>
<td>FIRST ACTION</td>
</tr>
<tr>
<td>21 INF DIV</td>
<td>LD 19</td>
<td>OCT 62</td>
<td></td>
<td>1200</td>
<td>80</td>
<td>95</td>
<td>90</td>
<td></td>
<td></td>
<td>ADVANCING</td>
<td>MOD RESISTENCE</td>
</tr>
<tr>
<td>20 INF DIV</td>
<td>KR 14</td>
<td>NOV 62</td>
<td></td>
<td>1500</td>
<td>85</td>
<td>80</td>
<td>90</td>
<td></td>
<td></td>
<td>ADVANCING</td>
<td>MOD RESISTENCE</td>
</tr>
<tr>
<td>19 INF DIV</td>
<td>LD 19</td>
<td>NOV 62</td>
<td></td>
<td>1400</td>
<td>75</td>
<td>80</td>
<td>80</td>
<td></td>
<td></td>
<td>ATTACKING</td>
<td>HIGH MORALE</td>
</tr>
<tr>
<td>22 INF DIV</td>
<td>TN 11</td>
<td>NOV 62</td>
<td></td>
<td>1500</td>
<td>95</td>
<td>75</td>
<td>85</td>
<td></td>
<td></td>
<td>SCREENING</td>
<td>HIGH MORALE</td>
</tr>
<tr>
<td>42 INF DIV</td>
<td>XL 21</td>
<td>JULY 62</td>
<td></td>
<td>1200</td>
<td>95</td>
<td>95</td>
<td>80</td>
<td></td>
<td></td>
<td>DEFENDING</td>
<td>MOD RESISTENCE</td>
</tr>
<tr>
<td>41 INF DIV</td>
<td>LD 19</td>
<td>AUG 62</td>
<td></td>
<td>1300</td>
<td>75</td>
<td>80</td>
<td>80</td>
<td></td>
<td></td>
<td>SCREENING</td>
<td>SWAMPY TERRN</td>
</tr>
<tr>
<td>24 INF DIV</td>
<td>PT 30</td>
<td>AUG 62</td>
<td></td>
<td>1300</td>
<td>95</td>
<td>75</td>
<td>95</td>
<td></td>
<td></td>
<td>SCREENING</td>
<td>COMBAT EXPER</td>
</tr>
</tbody>
</table>

**Figure 37.** Example of display studied by Ringel and Hammer (1964)
ditions. Obviously, all of the predicted times are significantly shorter than the actual times due to the complexity of the search task used by Ringel and Hammer. However, a regression that fits the actual times using the predicted times results in a reasonably high correlation \( r = .792 \). As with the Dodson and Shields study, that regression can be used to adjust the predicted times to take into account the greater complexity of the search task. Those adjusted times are also shown in Table 14.

The main effect of Number of Lines, which was significant in their study, is illustrated in Figure 38, which also shows the predicted (adjusted) search times averaged across the three levels of density. Clearly, the same order of conditions is exhibited by both the actual and predicted means.

Ringel and Hammer also found a very small, but significant, main effect of "density" of lines, as illustrated in Figure 39. The corresponding predicted (adjusted) search times, averaged across the three levels of amount of information, show essentially no effect, however. An interesting point is that if the density of the lines is increased even more— to a ratio of 1:1— then the predicted search time decreases in the same manner that one would extrapolate for the actual search times.

Gallan, Curran, and Lane (1977) studied a variety of CRT display formats for presenting Navy tactical information. Specifically, six formats varying from 6 to 40 display "items" were studied. An example of a 19-display-item format is shown in Figure 40. A display "item" was defined as the pairing of a 2-5 character mnemonic code (e.g., TN) with a 3-digit number. The subject's task
Figure 38. Actual and predicted search times for main effect of Number of Lines in Ringel and Hammer (1964) study
Figure 39. Actual and predicted search times for main effect of Density in Ringel and Hammer (1964) study
## Figure 40.

Example of a 19-display-item format studied by Callan, Curran, and Lane (1977)

<table>
<thead>
<tr>
<th>ENGA GE</th>
<th>AIR FRND INT/FGTR CAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>TN</td>
<td>392</td>
</tr>
<tr>
<td>PIF</td>
<td>977</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MK76</th>
<th>685</th>
<th>MK54</th>
<th>829</th>
<th>LB11A</th>
<th>317</th>
<th>MK84</th>
<th>563</th>
<th>FUEL</th>
<th>498</th>
</tr>
</thead>
<tbody>
<tr>
<td>MK44</td>
<td>244</td>
<td>FFAR</td>
<td>314</td>
<td>SU44A</td>
<td>822</td>
<td>MK58</td>
<td>164</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MK46</td>
<td>201</td>
<td>HVAR</td>
<td>885</td>
<td>N45</td>
<td>942</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MK57</td>
<td>246</td>
<td>MK52</td>
<td>578</td>
<td>NPS</td>
<td>406</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
was to search for a particular mnemonic code and then type the corresponding 3-digit number on a keyboard.

The actual and predicted search times for the six conditions listed in Table 14. As before, the predicted times are shorter than the actual times, but they are highly correlated ($r = .9$).

The difference between the actual and predicted times might be attributable to the time that it took the subjects in the Calla study to type the 3-digit number. Such typing time was not involved in Experiment 1 of the present study, from which the fictional system was derived. The adjusted times from the regression that fits the actual times using the predicted times are also in Table 14. The actual times and these adjusted times are plotted in Figure 41 as a function of number of display items. Calla et al. reported that their search times could be fit by the following linear function of number of display items ($r = .946$):

$$\text{Search Time} = .03458 \times \text{Number of Items} + 4.203$$

Interestingly, the predicted search times (based on overall display density, local density, number of groups, and size of groups) are better fit by such a function of number of display items ($r = .985$):

$$\text{Search Time} = .02887 \times \text{Number of Items} + 2.605$$

The main difference between these equations lies in their intercepts (4.203 vs. 2.605) rather than their slopes (.03458 vs. .02887), reinforcing the idea that the difference between the actual and predicted times is due to a constant effect, like
Figure 41. Actual and predicted search times from Callan, Curran, and Lane (1977)
the 3-digit number.

In addition to these studies that have reported empirical data on displays, several articles describing guidelines on display design have presented the author's concept of "good" and "bad" versions of the same display. One such article, Stewart (1976), has already been described (Section 3.5) and the two versions of the same basic display already presented (Figures 16 and 17). As shown in Table 16, the regression equations predict a search time of 4.76 sec and subjective rating of 3.09 for Stewart's "bad" version (Figure 16). Likewise, his "good" version (Figure 17) yielded a search time of 3.72 sec and subjective rating of 2.02. As one would hope, Stewart's redesign resulted in a significant improvement in both predicted search time (21.8% improvement) and subjective rating (45.7% improvement).

Marcus (1982) provided three examples of actual "before" and "after" displays from the redesign of a large information management system. These three pairs of displays are shown in Figures 42 through 44. Applying the regression equations to these displays resulted in the predicted search times and subjective ratings shown in Table 16. Interestingly, his redesigns had very little effect on the predicted search times. The redesigns shown in Figures 42 and 43 resulted in improvements in search time of only 1.8% and 6.0%, respectively, while the redesign in Figure 44 actually

9. Search time was predicted using overall density, local density, number of groups, and size of groups. Subjective rating was predicted using those four display measures plus number of items and item uncertainty.
<table>
<thead>
<tr>
<th>Reference</th>
<th>Predictions for &quot;Bad&quot; Version</th>
<th>Predictions for &quot;Good&quot; Version</th>
<th>% Improvement ((Bad-Good)/Bad)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Search Time</td>
<td>Subjective Rating</td>
<td>Search Time</td>
</tr>
<tr>
<td>Stewart (1976)</td>
<td>4.76</td>
<td>3.09</td>
<td>3.72</td>
</tr>
<tr>
<td>Marcus (1982)</td>
<td>4.32</td>
<td>2.10</td>
<td>4.24</td>
</tr>
<tr>
<td></td>
<td>4.47</td>
<td>2.26</td>
<td>4.20</td>
</tr>
<tr>
<td></td>
<td>4.17</td>
<td>2.17</td>
<td>4.30</td>
</tr>
<tr>
<td>Pakin &amp; Wray (1982)</td>
<td>4.79</td>
<td>3.80</td>
<td>3.56</td>
</tr>
<tr>
<td></td>
<td>3.56</td>
<td>1.45</td>
<td>25.7%</td>
</tr>
</tbody>
</table>
If you exit abnormally from SEEDIS, continue by typing
    restore
    seedis

Welcome to SEEDIS  VMS version 1.0
Type ? for expanded menus
Type $ before VMS commands
(HELP, REVIEW, SUBJECT, AREA, AGG, DISAGG,
 PROFILE, DATA, DISPLAY, BUGS, NETSTAT, SHOW, QUIT): subject
Please select both data and a geographic area.
FORTRAN STOP
WELCOME TO THE SYSTEM
YOU CAN ENTER

    EXPLAIN: TO SCAN KEYWORDS.
    SEARCH: TO LOCATE FILES CONTAINING KEYWORDS.
    QUIT: TO TERMINATE.

ENTER COMMAND OR ?COMMAND FOR MORE DETAIL.

---

WELCOME TO SEEDIS, VERSION 2.0
---

At any point in Seedis, you can type the following global
commands to get these services:

<table>
<thead>
<tr>
<th>Input</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>list and describe commands in this menu</td>
</tr>
<tr>
<td>help</td>
<td>describe the purpose of this menu’s commands</td>
</tr>
<tr>
<td>show</td>
<td>list and explain items to be selected</td>
</tr>
<tr>
<td>review</td>
<td>list current session status and history</td>
</tr>
<tr>
<td>cancel</td>
<td>delete current selections (depends upon context)</td>
</tr>
<tr>
<td>quit</td>
<td>return to previous menu</td>
</tr>
<tr>
<td>&quot;&lt;comment&gt;&quot;</td>
<td>enter a comment in Seedis log</td>
</tr>
</tbody>
</table>

shortly.

Please stand by. Your menu prompt will appear

SEEDIS: area, data, display, profile

---

Figure 42. First pair of ‘before’ (top) and ‘after’ (bottom)
displays presented by Marcus (1982)
TYPE ONE OF THE FOLLOWING COMMANDS...
?
HELP FOR THIS LIST OF COMMANDS
MORE FOR HOW TO GET HELP
TABLE TO SEE NEXT SCREENFULL
<N> FOR THE TABLE OF CONTENTS
* <COMMENT> TO ENTER A COMMENT IN THE LOG
DATA <SEQUENCE LETTERS> SELECT DATA CODES
CANCEL <SEQUENCE LETTERS> CANCEL DATA CODES
FOR X <C> SUBSTITUTE C FOR X IN DATA CODES
ALSO XX XXX XXXX Y YY YYYY YYYY
REVIEW LIST DATA SELECTIONS MADE SO FAR
SAVE SAVE DATA SELECTIONS AND RETURN
QUIT CANCEL DATA SELECTIONS AND RETURN
READY

<table>
<thead>
<tr>
<th>Input</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;line letter(s)&gt;</td>
<td>select one or more data elements by line letter</td>
</tr>
<tr>
<td>table</td>
<td>display table of contents for this database code</td>
</tr>
<tr>
<td>&lt;page number&gt;</td>
<td>display a particular page</td>
</tr>
<tr>
<td>CR</td>
<td>(carriage return) display the next page</td>
</tr>
<tr>
<td>?</td>
<td>list available commands in this menu</td>
</tr>
<tr>
<td>help</td>
<td>describe data element selection</td>
</tr>
<tr>
<td>show</td>
<td>display table of contents for this database</td>
</tr>
<tr>
<td>review</td>
<td>list current data element selections and history</td>
</tr>
<tr>
<td>cancel</td>
<td>delete current data element selections for this database</td>
</tr>
<tr>
<td>quit</td>
<td>return to database selection menu</td>
</tr>
</tbody>
</table>

DATA: <line letter(s)> , table, <page number> , CR

Figure 43. Second pair of 'before' (top) and 'after' (bottom) displays presented by Marcus (1982)
INTRODUCTION TO SEEDIS

The three major processes in SEEDIS are:

AREA: define a geographic study area (composed of states, counties, or census tracts)

DATA: select data appropriate to the geographic study area chosen. For example, for a study area consisting of a group of states, only state level data, and not county or tract level data, are appropriate.

DISPLAY: manipulate and display the data in table, chart, graph, and/or map form.

Normally AREA, DATA, and DISPLAY are performed in the order given. However, once the geographic study area is defined (AREA), one may alternate between DISPLAY and the selection and extraction of additional items in DATA.

TYPE MORE TO SEE NEXT SCREENFULL
TYPE ? FOR A LIST OF COMMANDS

---

SEEDIS: area, data, display, profile

USING SEEDIS

LBL's Seedis is an experimental information system that includes integrated program modules for retrieving, analyzing, and displaying selected portions of geographically linked databases. Program modules in Seedis include:

area
select geographic area (level and scope of analysis)
data
select, extract, enter, or transform data
display
manipulate and display data in tables, maps, and charts
profile
produce standard socio-economic reports for selected area

Normally, Area, Data, and Display are used in the order given. However, once the geographic study area is defined in Area, you may alternate between Display and the selection, extraction, or entering of additional items in Data.

SEEDIS: area, data, display, profile

---

Figure 44. Third pair of 'before' (top) and 'after' (bottom) displays presented by Marcus (1982)
resulted in a small detriment to search time of 3.1%. However, Marcus' redesigns resulted in substantial improvements in predicted subjective ratings, ranging from 49.6% to 56.7%. These large improvements in predicted subjective ratings are not surprising since Marcus primarily focused on aligning data elements by using a small number of tab stops. This resulted in substantial reductions in the "item uncertainty" measure (e.g., from 7.63 bits to 5.24 bits for Figure 44), which is one of the most important factors in the prediction of subjective rating.

Fakin and Wray (1982) presented "good" and "bad" versions of a display for a hypothetical hotel reservation and billing system. The "bad" version is shown in Figure 45. They claimed that this display is too cluttered and should be broken into two separate displays: "It is better to have two neat, well-formatted screens with some unused space than one cluttered screen with no unused space" (Fakin & Wray, 1982, p. 37). The two new displays are shown in Figures 46 and 47. Applying the prediction system to all three displays results in the predicted search times and subjective ratings shown in Table 16. Each of the two new displays is, by itself, a substantial improvement over the old display, both in search time (25.7% improvement in both cases) and in subjective rating (44.2% and 61.8% improvements).

The interesting question this raises is whether these predictions justify separating the one screen into two. These data by themselves cannot provide the answer to that question, but they can help. The predicted search time for both redesigned displays was 1.23 sec less than that for the original display. Since the data
### Figure 45. 'Bad' version of display for a hypothetical hotel reservation and billing system from Pakin and Wray (1982)
Figure 46. First of two new displays presented by Pakin and Wray (1982) to replace the display in Figure 45

Figure 47. Second of two new displays presented by Pakin and Wray (1982) to replace the display in Figure 45
are now contained in two displays rather than one, however, one must consider the extra time that may be required to decide which display is needed and to retrieve that display. If that total extra time is greater than 1.23 sec then the split into two screens would not be justified.

One can easily envision that these hotel reservation and billing screens would be just two of many screens in a complete hotel management system. It is quite likely than all these screens would be selected from a single "master menu." The comparison to make, then, is between the time it takes to choose a single menu item called, perhaps, "RESERVATION AND BILLING" and the time it takes to choose between two menu items called, perhaps, "RESERVATION" and "BILLING." In this particular case it seems clear that the split into two screens is justified since it certainly could not take more than 1.23 sec longer to choose between the two items called "RESERVATION" and "BILLING" than to choose the one item "RESERVATION AND BILLING." In cases where the functional basis of the split is not so obvious, however, the split may not be justifiable. In any event, the prediction system provides information that is useful in making that decision.

In summary, the prediction system seems to have performed well when applied to displays in the literature. For those displays that have associated empirical data (Table 14), the correlations between actual and predicted search times were generally high, ranging from .792 to .985. Likewise, for those displays that various authors presented as "good" and "bad" ways of presenting the same data (Table 16), the predicted subjective ratings were always
better for the "good" formats, and the predicted search times were always better or essentially the same.
7. CONCLUSIONS

The results of Experiments 1 and 2 and the results of applying the prediction system to displays in the literature all support the validity of the system. Given a set of alternative formats for presenting some alphanumeric data, the prediction system can quite accurately predict the relative search times and subjective ratings for the formats.

It is important to note, however, that this prediction system is intended to be used for making comparisons between alternative formats for presenting the same basic data. For example, it is not really meaningful to compare Pakin and Wray's (1982) redesigned hotel reservation display (Figure 46) to one of Marcus' (1982) redesigned displays (e.g., Figure 44). To say that Figure 46 is a better display than Figure 44 because of the difference in predicted search times (3.56 sec vs. 4.30 sec) is inappropriate because the displays were designed for quite different uses.

Given that the prediction system is valid for relative comparisons, it seems appropriate to study it once more to see what can be learned about the design of alphanumeric displays. The following conclusions can be drawn:

1. A display that optimizes search times is not necessarily a display that optimizes subjective ratings, and vice versa.

This is based on the fact that different display parameters and coefficients are involved in the regression equations for predicting search time and subjective rating. Thus, a system designer may need to decide which to optimize, or how to
assign relative weights to the predictions in deciding which format to use. If equal weighting is desired, the sum of z-score transformations may be appropriate. (See Section 5.4.)

2. The two most important display variables in determining search time appear to be those associated with the grouping of characters.

This is based on the results of Experiment 1, which found that "number of groups" and "size of groups" provided the best-fitting two-variable regression equation ($R^2 = .418$), and Experiment 2, which found that the same regression equation generalized well to a new set of displays ($R^2 = .508$). Since both of these variables have positive coefficients in the regression, if either value increases then the search time increases.

For example, consider the displays shown in Figures 48 through 50, and their corresponding display characteristics shown in Table 17. Given that a fixed amount of data is to be presented (as is the case with Figures 48 through 50), there is generally a reciprocal relationship between number and size of groups: one can present the data in one large group as in Figure 48, many small groups as in Figure 49, or something in between as in Figure 50. The regression equation predicts that the best solution is one which minimizes both number of groups and size of groups. Thus, the predicted search times (using the four-variable model) for Figures 48-50 are 4.87 sec, 6.74 sec, and 2.84 sec, respec-
Figure 48. Display with one large group of characters

Figure 49. Display with 240 small groups of characters
### Figure 50. Display with 32 intermediate-size groups of characters
TABLE 17. Values of Six Display Variables, Predicted Search Time, and Predicted Subjective Rating for Five Prototypical Displays

<table>
<thead>
<tr>
<th>Display</th>
<th>Overall Density</th>
<th>Local Density</th>
<th># of Groups</th>
<th># of Grp. Size</th>
<th># of Items</th>
<th>Uncertainty</th>
<th>Search Time</th>
<th>Subjective Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 48</td>
<td>25.0%</td>
<td>84.7%</td>
<td>1</td>
<td>14.0°</td>
<td>12</td>
<td>3.58 bits</td>
<td>4.87</td>
<td>2.00</td>
</tr>
<tr>
<td>Fig. 49</td>
<td>25.0%</td>
<td>24.9%</td>
<td>240</td>
<td>0.6°</td>
<td>240</td>
<td>7.91 bits</td>
<td>6.74</td>
<td>8.51</td>
</tr>
<tr>
<td>Fig. 50</td>
<td>25.0%</td>
<td>31.4%</td>
<td>32</td>
<td>1.7°</td>
<td>96</td>
<td>6.58 bits</td>
<td>2.84</td>
<td>1.76</td>
</tr>
<tr>
<td>Fig. 51</td>
<td>25.0%</td>
<td>50.7%</td>
<td>32</td>
<td>1.7°</td>
<td>96</td>
<td>6.58 bits</td>
<td>3.40</td>
<td>2.77</td>
</tr>
<tr>
<td>Fig. 52</td>
<td>25.0%</td>
<td>30.9%</td>
<td>32</td>
<td>2.2°</td>
<td>96</td>
<td>9.14 bits</td>
<td>2.88</td>
<td>2.83</td>
</tr>
</tbody>
</table>
tively. In general, then, it is better to present data in an intermediate number of medium-sized groups (as in Figure 50) than either one large group or many small groups.

3. The two most important display variables in determining the subjective rating of a display appear to be "local density" and "item uncertainty."

This is based on Experiment 1, which found that those two variables provided the best-fitting two-variable regression equation ($R^2 = .682$), and Experiment 2, which found that the same regression equation generalized well to a new set of displays ($R^2 = .539$). As with the two grouping variables, both of these variables have positive coefficients in the regression equation, so if either value increases the subjective rating increases (worsens). Unlike the grouping measures, however, local density and item uncertainty are not clearly related to each other. Local density is essentially a measure of how "tightly packed" the display is. For example, Figure 51 is a more tightly packed version of Figure 50. As Table 17 shows, the only display variable that differs between the two figures is local density: 31.4% for Figure 50 vs. 50.7% for Figure 51. As a result, the predicted subjective rating (using all six display variables) is worse for Figure 51 (2.77) than for Figure 50 (1.76). Item uncertainty, on the other hand, is a measure of how "poorly aligned" the items on the display are. For example, Figure 52 is a poorly aligned version of Figure 50. As Table 17
Figure 51. Tightly packed version of Figure 50

Figure 52. Poorly aligned version of Figure 50
shows, the only display variable that differs significantly between the two figures is item uncertainty: 6.58 bits for Figure 50 vs. 9.14 bits for Figure 52. As a result, the predicted subjective rating is worse for Figure 52 (2.83) than for Figure 50 (1.76).

In conclusion, a computer-based tool for objectively evaluating the usability of a display has been developed and validated. It should be particularly useful to system designers who must choose between alternative designs for a user interface, but who do not have the time or resources to collect human performance data.
REFERENCES


Williams, C. M. Horizontal versus vertical display of numbers. *Human Factors*, 1966, 8, 237-238.


This program measures a variety of objective characteristics of "formatted" alphanumeric displays (e.g., CRT displays) for use in predicting how well a user would be able to extract information from the displays. The characteristics measured are:

1. Overall density - the number of characters on the display, expressed as a percentage of the character spaces available.

2. Local density - the number of other characters within a 5-deg (visual angle) circle centered on each character, expressed as a percentage of the number of character spaces within that circle. Essentially, this is a measure of how "tightly packed" the display is.

3. Number of groups - the number of groups of characters detected using a proximity clustering technique.

4. Average size of groups - the average visual angle subtended by the groups of characters, weighted by the number of characters in each group.

5. Number of items - the number of distinct labels or data items on the display. This is highly correlated with overall density.

6. Layout complexity - a measure of the amount of information conveyed simply by the layout of items on the display. Using vertical alignments of items decreases layout complexity.

7. Consistency - a measure of how well-correlated the spatial arrays of characters on a set of displays are with each other.

Input to this program is a file (or files) that contains a literal example of the display to be evaluated.

NOTE: In order to compile and link this program, it must be linked with the math library. Use the following to compile on UNIX:

```
cc display.c -lm
```
else if (strcmp(argument,"-nocheck")==0)
    nocheck=1;
else if (strcmp(argument,"-noorder")==0)
    noorder=1;
else if (strcmp(argument,"-noconsist")==0)
    noconsist=1;
else if (strcmp(argument,"-nostatus")==0)
    nstatus=1;
else if (strcmp(argument,"-append")==0){
    append=1;
    fpsumm = fopen("summ.data","a");
}
else
    fprintf(stderr,"Unknown option: %s\n", argument);
continue; /* Get next argument- this was option*/
}
fp = fopen("argv", "r");
if (fp==NULL) {
    fprintf(stderr,"Cannot open %s\n","argv");
    continue;
} else {
++n_displays;
filename[n_displays] = *argv;
++no_displays;
if (nstatus==0)
    fprintf(stderr,"Analyzing %s\n","argv");
printf("File %s: \n","argv");
cputchar (get_mask(fp));
cputchar (get_mask(fp));
if (count<2) {
    fprintf(stderr,"File %s contains fewer than 2 characters\n","argv");
    fprintf(stderr,"Further analyses would not be meaningful\n");
    continue;
}
n_char=count;
total_spaces= n_cols * n_rows;
overall_dens= n_char/total_spaces;
for_dens[n_displays] = overall_dens*100;
if (nstatus==0) {
    density();
loc_dens[n_displays] = local_dens*100;
}
if (nogroup==0) {
    find_neighbors();
n_groups = find_groups(mask,count);
if (nocheck==0)
    n_groups = check_groups(n_groups,count);
groups[n_displays] = n_groups;
print_groups(n_char);
putdashes();
grp_percent(n_groups,n_char);
grp_pct[n_displays] = pct_grp;
}
if (noorder==0) {
    n_items = find_items();
    no_items[n_displays]=n_items;
    order1[n_displays] = orderliness(n_items);
}
if (noprint==0)
    overlay_mask();
fclose(fp);
if (append==1) {
    i = n_displays;
    fp = fopen(filename[i],"a");
    fprintf(fp,"\n");
    fprintf(fp," %7.1f",ov_dens[i]);
    fprintf(fp," %7.1f",loc_dens[i]);
    fprintf(fp," %7.0f",groups[i]);
    fprintf(fp," %7.1f",grp_pct[i]);
    fprintf(fp," %7.0f",no_items[i]);
    fprintf(fp," %7.2f\n",order1[i]);
c=12;
   putc(c,fp);
    fclose(fp);
    /* Now add data to summary file */
    fp = fsumm;
    fprintf(fp," %7.1f",ov_dens[i]);
    fprintf(fp," %7.1f",loc_dens[i]);
    fprintf(fp," %7.0f",groups[i]);
    fprintf(fp," %7.1f",grp_pct[i]);
    fprintf(fp," %7.0f",no_items[i]);
    fprintf(fp," %7.2f\n",order1[i]);
}
/* Print summary table for this display */
i=n_displays;
printf(" |	Overall		Local		Av. Visual		Layout\n"):
printf("Filename	Density	Density	#Grps	Angle of Grps	#Items	Complexity\n"):
printf("%-14s",filename[i]);
printf("%4.1f%%", ov_dens[i]);
printf("%8.1f%%", loc_dens[i]);
printf("%7.0f", groups[i]);
printf("%10.1f deg", grp_pct[i]);
printf("%9.0f", no_items[i]);
printf("%9.2f\n", order1[i]);
i=12; putchar(1); /*form feed*/
exit(0);

if (noconsist==0)
    consist=consistency(n_displays);

/* Now calculate summary statistics */

count = n_displays+1;
for (i=1; i<=n_displays; ++i) {
    ov_dens[count] = ov_dens[count] + ov_dens[i];
    loc_dens[count] = loc_dens[count] + loc_dens[i];
    groups[count] = groups[count] + groups[i];
    grp_pct[count] = grp_pct[count] + grp_pct[i];
    no_items[count] = no_items[count] + no_items[i];
    order11[count] = order11[count] + order11[i];
}

i = n_displays;
ov_dens[count] = ov_dens[count]/i;
loc_dens[count] = loc_dens[count]/i;
groups[count] = groups[count]/i;
grp_pct[count] = grp_pct[count]/i;
no_items[count] = no_items[count]/i;
order11[count] = order11[count]/i;

printf("SUMMARY:\n\n");
printf("Overall Local Av. Visual Layout\n");
printf("Filename Density Density #Grps Angle of Grps #Items Complex. Consist.\n\n");
if (n_displays==1)
    count=1; /* Don't print averages */
for (i=1; i<count; ++i) {
    if (i<n_displays) /* then it's average data */
        printf("\nAverages\n");
    else
        printf("%s",filename[i]);
    printf("%.4f\n", ov_dens[i]);
    printf("%.8f\n", loc_dens[i]);
    printf("%.7f\n", groups[i]);
    printf("%.10f\n", grp_pct[i]);
    printf("%.9f\n", no_items[i]);
    printf("%.9.2f\n", order11[i]);
    if (i<n_displays)
        printf("%s\n",consist);
    else
        printf("\n");
}

get_specs() /* get CRT specifications */
int maxrows, maxcols;
double get_float();
maxrows=MAXROW;
maxcols=MAXCOL;
n_rows = n_cols = 0;
char_height = char_width = 0.0;
fprintf(stderr,"Please enter the characteristics of the display device:
");
fprintf(stderr," Maximum number of characters per line? ");
n_cols = get_int(maxcols);
fprintf(stderr," Maximum number of lines? ");
n_rows = get_int(maxrows);
fprintf(stderr," Distance between character centers horizontally? ");
char_width = get_float();
fprintf(stderr," Distance between character centers vertically? ");
char_height = get_float();

get_int(x)
int x;
{
    int n, c;
    while ((n=scanf("%d", &c)) != 1 || c < 1) {
        fprintf(stderr," ERROR: Must be >0 and <%d\n ?", x);
        return(c);
    }

double get_float()
{
    int n;
    double c;
    while ((n=scanf("%f", &c)) != 1 || c <= 0) {
        fprintf(stderr," ERROR: Must be >0\n ?");
        return(c);
    }

get_mask(fp)                                 /* put mask into array */
FILE *fp:
{
    int col, row, count, c, tab;
    for (row=0; row<n_rows; ++row) {
        for (col=0; col<n_cols; ++col) {
            mask[row][col] = -1; /* initialize mask to all -1 */
            orig_mask[row][col] = 32; /* make this copy all spaces */
        }
    }

    col=0;
    row=1;
    count=0;
    while ((c=getc(fp)) != EOF && c != 126) {
        ++col;
        if (c==9) { /* it's a tab, need to expand to spaces */

tab=1;
while (tab<=col) /*find position to tab to */
    tab = tab + TAB;
while (col<tab) /* now move to it */
    ++col;
    --col; /*back up 1, since col will be incremented before next char*/
}
    if (c>32 & c<127) { /* it's a character */
        ++count;
        mask[row][col] = 0; /* set char positions to 0 */
        orig_mask[row][col] = c;
    }
    if (c=='\n' || col > n_cols)
        ++row;
        col=0;
    while (c!='\n')
        c=getc(fp); /* read to begin of next line*/
    putchar(c);
    if (row > n_rows)
        break;
}
return(count);

get_term() /* get terminal characteristics */
{
    double dist,view_dist;
    int x, y, counter;
    FILE *fp, *fopen();
    char term[20];
    fprintf(stderr,"Delimiter between labels and data (1 character)? ");
    scanf("%c", &delimiter);
    fprintf(stderr,"Coefficient to use in grouping algorithm? ");
    scanf("%f", &mean_coeff);
    fprintf(stderr,"Viewing distance from eye to display? ");
    scanf("%f", &view_dist);
    visual_angle= TANGENT * view_dist;
    fprintf(stderr,"Type of terminal? ");
    scanf("%s", term);
    while ((fp=fopen(term,"r")) == NULL) {
        fprintf(stderr,"Terminal type %s not known\n", term);
        get_specs();
        build_distances(term);}
    counter=0; n_rows=0; n_cols=0;
    while (fscanf(fp, INFORMAT, &dist, &x, &y) != EOF) { /* read distances */
        if (x==1 & y==0)
            char_width=dist;
        if (x==0 & y==1)
char_height=dist;
distances[counter]=dist;
search_order[counter][0]=x;
search_order[counter][1]=y;
if (x > n_cols)
    n_cols = x;
if (y > n_rows)
    n_rows = y;
++counter;
}

build_distances(term)
char term[];
{
    FILE *fp, *fp1, *fopen();
    int x,y;
    double dist,a,b;
    fp= fopen("term.tmp","w");
    fprintf(stderr,"Calculating distances\n");
    for (y=0; y<n_rows; ++y) {
        for (x=0; x<n_cols; ++x) {
            a = x*char_width;
            b = y*char_height;
            dist = sqrt(a*a + b*b);
            fprintf(fp,OUTFORMAT,dist,x,y);
        }
    }
    fclose(fp);
    fprintf(stderr,"Sorting distances\n");
    system("sort -n term.tmp > sorted.tmp");
    fp = fopen("sorted.tmp","r");
    fp1= fopen(term,"w");
    while (fscanf(fp,INFORMAT,&dist,&x,&y) != EOF) {
        fprintf(fp1,OUTFORMAT,dist,x,y);
    }
    fclose(fp);
    fclose(fp1);
    system("rm term.tmp sorted.tmp");
}

find_neighbors()
{
    int n_char,row,col,counter,c,x,y,c1,c2,c3,c4,a1,a2,b1,b2;
    double dist, sum_distances, ss_distances, mean, std_dev;
    sum_distances = 0; ss_distances = 0; n_char = 0;
    for (row=1; row<n_rows; ++row) {
        for (col=1; col<n_cols; ++col) {
            if (mask[row][col]==0) { /* it's a character */
                ++n_char;
                dist = distances[counter];
                fprintf(fp,OUTFORMAT,dist,x,y);
            }
        }
c=(-1); counter=0;
while (c<0) { /* search for neighbor */
  ++counter;
  x=search_order[counter][0];
  y=search_order[counter][1];
  a1=rows-y; a2=rows+x;
  b1=cols+x; b2=cols+x;
  c1= (a1<1 ) ? (-1) : mask[a1][b1];
  c2= (a1<1 ) ? (-1) : mask[a1][b2];
  c3= (a2>cols ) ? (-1) : mask[a2][b1];
  c4= (a2>cols ) ? (-1) : mask[a2][b2];
  c= max(c1,c2,c3,c4);
  if (dist = distances[counter])
    sum_distances = sum_distances + dist;
    ss_distances = ss_distances + dist*dist;
  }
}
mean= sum_distances/n_char;
std_dev = (ss_distances - (sum_distances * sum_distances)/n_char);
threshold = mean_coeff * mean + sd_coeff * std_dev;

max(c1,c2,c3,c4)
int c1,c2,c3,c4;
{
    int c;
    c = (c1>c2) ? c1 : c2;
    c = (c>c3) ? c : c3;
    c = (c>c4) ? c : c4;
    return(c);
}

find_groups(my_mask,n_char)
int my_mask[MAXROW][MAXCOL];
int n_char;
{
    int grp_nbr,n_groups,row,col,counter,n,m,x,y,a[3],b[3],c,i,j,n_hit;
    grp_nbr=0 ; n_groups=0;
    for (row=1;row<=rows;++row)
        for (col=1;col<=cols;++col) {
            if (my_mask[row][col] > (-1)) { /* it's a char */
                --n_char;
                n_hit=0;
                if (my_mask[row][col] == 0) { /* new group */
                    ++n_groups;
                    grp_nbr= n_groups;
                }
                /* search for neighbor */
                ++counter;
                x=my_mask[row][col];
                y=my_mask[row][col];
                a1=rows-y; a2=rows+x;
                b1=cols+x; b2=cols+x;
                c1= (a1<1 ) ? (-1) : mask[a1][b1];
                c2= (a1<1 ) ? (-1) : mask[a1][b2];
                c3= (a2>cols ) ? (-1) : mask[a2][b1];
                c4= (a2>cols ) ? (-1) : mask[a2][b2];
                c= max(c1,c2,c3,c4);
                if (dist = distances[counter])
                    sum_distances = sum_distances + dist;
                    ss_distances = ss_distances + dist*dist;
                }
            }
        }
    mean= sum_distances/n_char;
    std_dev = (ss_distances - (sum_distances * sum_distances)/n_char);
    threshold = mean_coeff * mean + sd_coeff * std_dev;
    }
    max(c1,c2,c3,c4)
    int c1,c2,c3,c4;
    {
        int c;
        c = (c1>c2) ? c1 : c2;
        c = (c>c3) ? c : c3;
        c = (c>c4) ? c : c4;
        return(c);
    }
}
if (my_mask[row][col] > 0) /* old group */
grp_nbr = my_mask[row][col];
counter = 0;
while (distances[counter] <= threshold) {
    x = search_order[counter][0];
    y = search_order[counter][1];
    if (y==0)
        n=1;
    else
        n=2;
    if (x==0)
        m=1;
    else
        m=2;
    for (i=1;i<=n;++i)
        for (j=1;j<=m;++j)
            if (a[i]<1 || a[i]>n_rows || b[j]<1 || b[j]>n_cols)
                c=(-1);
            else
                c=my_mask[a[i]][b[j]];
    if (c>(-1)) {
        if (counter>0)
            ++n_hit;
        if (c==0)
            c=grp_nbr;
        my_mask[a[i]][b[j]] = grp_nbr;
    }
    backtrack(&grp_nbr,&n_groups,c,my_mask);
    counter++;
}
}
if (n_char==0)
    break;
}
if (n_char==0)
    break;
return(n_groups);
int smallest, good_split, l, large_small, use_split;

for (group=1; group<=n_groups; ++group) {/* counter for groups */
    n_in_group = get_group(grp_mask.group);
    if (n_in_group > COEFF*n_char) {/* candidate for splitting */
        n_splits = find_splits(grp_mask.n_in_group.splits);
        for (c=1; c<n_splits; ++c) {
            row=spitits[c][1];
            col=spits[c][2];
            grp_mask[row][col]=(-2);
        }
        /*Now see if taking out ALL split char’s causes >1 grp*/
        new_n_groups = find_groups(grp_mask,(n_in_group-n_splits));
        if (new_n_groups > 1) {/*Taking all out did cause a split*/
            good_split = 0;
            large_small = 0;
            for (c=1; c<n_splits; ++c) {/*Take out 1 at a time*/
                row = splits[c][1];
                col = splits[c][2];
                grp_mask[row][col] = (-2);
                new_n_groups = find_groups(grp_mask,(n_in_group-1));
                if (new_n_groups > 1) {/*Caused a split*/
                    /*Make sure no new grp is too small*/
                    count(grp_mask,n_each_group);
                    smallest = MAX;
                    for (i=1; i<=new_n_groups;++i){
                        new=n_each_group[i];
                        if (new<smallest)
                            smallest=new;
                    }
                    if (smallest>MN) {
                        good_split=1;
                        if (smallest>large_small) {
                            large_small=smallest;
                            use_split=c;
                        }
                    }
                }
            }
        }
    } /*Now update main mask with grp no’s for split-off groups -- give any split-off
group a group no. higher than the last grp*/
for (row=1; row<=n_rows; ++row) {
    for (col=1; col<=n_cols; ++col) {
        c=mask[row][col];
        if (c==group){
            if (grp_mask[row][col]>1)
                mask[row][col]=grp_mask[row][col]+n_groups-1;
        }
    }
}

n_groups=n_groups + new_n_groups -1;
--group; /*Check what's left of this
    group again to make sure
    it's not still too big*/

return(n_groups);
}

group(grp_mask, group)
int grp_mask[MAXROW][MAXCOL], group;
{
    int n_in_group, row, col;
    n_in_group=0;
    for (row=1; row<=n_rows; ++row) {
        for (col=1; col<=n_cols; ++col) {
            if (mask[row][col]==group) {
                ++n_in_group;
                grp_mask[row][col]=0;
            } else
                grp_mask[row][col]=(-1);
        }
    }
    return(n_in_group);
}

count(my_mask, n_each_group)
int my_mask[MAXROW][MAXCOL];
int n_each_group[200];
{
    int row, col, i;
    for (i=0; i<100; ++i)
        n_each_group[i] = 0;
    for (row=1; row<=n_rows; ++row) {
        for (col=1; col<=n_cols; ++col) {
            i = my_mask[row][col];
            n_each_group[i] += 1;
        }
    }
}
if (i>(-1))
    ++n_each_group[i];
}
}

find_splits(grp_mask,n_in_group,splits)
int grp_mask[MAXROW][MAXCOL], n_in_group, splits[2C][3];
{
    double total, mean:
    int n_surround[MAX][3], row, col, max_splits, n_char, c, counter;
    int n_splits, x, y, a[3], b[3], n_hit, i, j, m, n, max, check:

    /* First find how many char's are in thresh. dist of each char*/
    n_char=0;
    for (row=1; row<=n_rows; ++row) {
        for (col=1; col<=n_cols; ++col) {
            if (grp_mask[row][col] > (-1)) { /*it's a char */
                n_char=0;
                ++n_char;
                counter=0;
                while (distances[counter] <= threshold) {
                    x = search_order[counter][0];
                    y = search_order[counter][1];
                    if (y==0)
                        n=1;
                    else
                        n=2;
                    if (x==0)
                        m=1;
                    else
                        m=2;
                    for (i=1;i<=n;++i){
                        for (j=1;j<=m;++j){
                            if (a[i]<1 || a[i]>n_rows || b[j]<1 || b[j]>n_cols)
                                c=(*-1);
                            else
                                c=mask[a[i]][b[j]];
                            if (c>(-1)) {
                                if (counter>0)
                                    ++n_hit;
                            }
                        }
                    }
                }
                n_surround[n_char][0]=n_hit;
                n_surround[n_char][1]=row;
n_surround[n_char][2]=col;
}
    if (n_char==n_in_group)
        break;
    if (n_char==n_in_group)
        break;

/* Get mean no. of char's surrounding each char */
    n_hit=0;
    for (i=1; i<=n_in_group; ++i)
        n_hit=n_hit+n_surround[i][0];
    total=n_hit;
    mean = total/n_in_group;
    mean = round((.75)*mean);
    max=5; /*max. no. of surrounding char's a char can have & be a split candidate*/
    check= (mean < max) ? mean : max;

/* Find the char's that are candidates for splits */
    n_splits=0;
    max_splits=10; /*max. no. of split char's to consider */
    for (i=1; i<=check; ++i) {
        for (j=1; j<=n_in_group; ++j) {
            if (n_surround[i][0]==1) {
                ++n_splits;
                for (n=0; n<2; ++n)
                    splits[n_splits][n]=n_surround[j][n];
            }
        }
    }
    if (n_splits>=max_splits)
        break;
    if (n_splits>=max_splits)
        break;
return(n_splits);

grp_percent(n_groups, n_char)  /* Calculate av. visual angle of groups */
int n_groups;
double n_char;
{
    int x, y, l, j, a[3], b[3], c, counter;
    int row, col, group, n_in_group, n_hit;
    double min_col, max_col, min_row, max_row, total, view_dist, dist;
    double count;
    total = 0;
    for (group=1; group<=n_groups; ++group) { /*find min & max grp members*/
        n_in_group = 0;

min_col = n_cols; max_col = 0;
min_row = n_rows; max_row = 0;
for (row=1; row<=n_rows; ++row) {
  for (col=1; col<=n_cols; ++col) {
    if (mask[row][col]==group) {
      ++n_in_group;
      if (col<min_col)
        min_col=col;
      if (col>max_col)
        max_col=col;
      if (row<min_row)
        min_row=row;
      if (row>max_row)
        max_row=row;
    }
  }
}

view_dist = visual_angle/TANGENT;
col = round((min_col+max_col)/2); /*center col & row of grp*/
row = round((min_row+max_row)/2);
counter = 0;
n_hit = 0;
while (n_hit != n_in_group) {
  x=search_order[counter][0];
  y=search_order[counter][1];
  for (i=1; i<=2; ++i) {
    for (j=1; j<=2; ++j) {
      if (a[i]<1 || a[i]>n_rows)
        c=1;
      else if (b[j]<1 || b[j]>n_cols)
        c=1;
      else
        c=mask[a[i]][b[j]];
      if (c==group) { /* hit a char in grp */
        ++n_hit;
        mask[a[i]][b[j]] = 0; /* don't want to count again */
      }
    }
  }
  ++counter;
pct_grp = total/n_char;
return(pct_grp);
}

print_groups(n_char)
double n_char;
{
    int row, col, c, base;
    double n_printed;
    base = 48;
    n_printed = 0;
    for (row=1; row<n_rows; ++row)
    {
        for (col=1; col<n_cols; ++col)
        {
            c = base + mask[row][col];
            if (c==46) /* split character */
                c = 126;
            else if (c<48)
                c = 32; /* space */
            while (c>126)
                c = c-94;
            putchar(c);
            if (c>=32)
                ++n_printed;
        }
    }
    if (n_printed==n_char) /* don't just print blank lines */
        printf("\n");
    break;
}

putchar(\n);

putedashes()
{
    int i;
    for (i=1; i<=80; ++i)
        putchar('-');
    putchar(\n);
}
for (row=1; row<=n_rows; ++row) {
    for (col=1; col<=n_cols; ++col) {
        if (my_mask[row][col]==change)
            my_mask[row][col]=use;
        if (my_mask[row][col]>change)
            --my_mask[row][col];
    }
}

for (grp_nbr=use;
    ++p_n_groups;
} density()
{
    int row,col,n_char,counter,x,y,1,2,3,c[3],b[3];
    double value, sum_values, sum_weights;
    n_char=0; sum_values=0; sum_weights=0;
    for (row=1; row<=n_rows; ++row) {
        for (col=1; col<=n_cols; ++col) {
            if (mask[row][col] > (-1)) { /* it's a char */
                ++n_char;
                counter=1;
                while (distances[counter]<=visual_angle) {
                    x=search_order[counter][0];
                    y=search_order[counter][1];
                    for (i=1; i<=2; ++i) {
                        for (j=1; j<=2; ++j) {
                            if (a[i]<1 || a[i]>n_rows)
                                c=(-1);
                            else if (b[j]<1 || b[j]>n_cols)
                                c=(-1);
                            else
                                c=mask[a[i]][b[j]];
                            value=((10)/visual_angle)*distances[counter]+10;
                            if (c>=(-1)) /*hit surrounding char*/
                                sum_values=sum_values+value;
                            sum_weights = sum_weights + value;
                        }
                    }
                    ++counter;
                }
            }
        }
    }
    local_dens = sum_values/sum_weights;
find_items() /* find distinct labels and data items */
{
    int row, col, c, n_items, prev_c, have_start;
    n_items=0;
    for (row=0; row<n_rows; ++row)
        for (col=0; col<n_cols; ++col)
            have_start=0;
    c=orig_mask[row][col];
    if (c>=48 && c<=57) { /* it's a digit */
        ++n_items;
        item_positions[n_items][0] = row; /* assume for now, since */
        item_positions[n_items][1] = col; /* it could be alphanumeric. */
        prev_c = 0;
        while (++col<n_cols) { /* read to end of item */
            c = orig_mask[row][col];
            if (c==46 || c==32 || c==58) { /* decimal pt, space, or colon (as in time) */
                if (have_start==0) {
                    have_start=1;
                    item_positions[n_items][0] = row;
                    item_positions[n_items][1] = col;
                }
            }
            if (c<48 || c>57) /* alpha char */
                have_start = 1;
            if (c==32 && prev_c==32) /* 2 spaces in a row */
                break;
            prev_c = c;
        }
    } else if (c>32 && c<125) { /* it's a char */
        ++n_items;
        item_positions[n_items][0] = row;
        item_positions[n_items][1] = col;
        prev_c = 0;
        while (++col<n_cols) { /* read to end of item */
            c = orig_mask[row][col];
            if (c==32 && prev_c==32) /* 2 spaces in a row */
                break;
            if (c==delimiter) /* label-data delimiter */
                break;
            prev_c = c;
        }
    }
    return(n_items);
}

double orderliness(no_items) /* Calculate orderliness using information theory */
int no_items;
{
    int i, j, n_classes;
    double starts[MAXCOL+1][2], n_items;
    double order[2], prob, uncertainty, sum_uncertainty, total;

    n_items = no_items;
    for (i=1; i<=MAXCOL; ++i) {
        for (j=0; j<i; ++j) {
            starts[i][j] = 0;
        }
    }
    for (i=1; i<=n_items; ++i) {
        for (j=0; j<=i; ++j) {
            ++starts[item_positions[i][j]][j];
        }
    }
    for (j=0; j<=i; ++j) { /* 0 is vert. complexity, 1 is horizontal */
        sum_uncertainty = 0;
        n_classes = 0;
        for (i=1; i<=MAXCOL; ++i) {
            if (starts[i][j] > 0) {
                ++n_classes; /*count how many unique distances*/
                prob = starts[i][j]/n_items;
                uncertainty = (-prob)*(log10(prob)*3.32193);
                sum_uncertainty = sum_uncertainty + uncertainty;
            }
        }
        order[j] = sum_uncertainty;
    }
    total = order[0] * order[1];
    return(total);
}

round(x)    /* round a double to an int */
double x;
{
    int y, y1, a;
    double x1;
    x1 = x + (.5);
    y1 = x1; /* truncates x1 to integer portion */
    y = x;   /* truncates x to integer portion */
    if (y==y1)
        a=y;
    else
        a=y+1;
    return(a);
}
overlay_mask()
{
    int row, col, c;
    for (row=0; row<n_rows; ++row) {
        for (col=0; col<n_cols; ++col) {
            c = orig_mask[row][col];
            if (c>32 && c<127) /* it's a character */
                ++cum_mask[row][col];
        }
    }
}

double consistency(no_displays)
int no_displays;
{
    int row, col, n;
    double n_displays, consist;
    double n_occupied, sum;
    n_displays = no_displays;
    n_occupied = 0.0; sum = 0.0;
    if (n_displays < 2)
        printf("Cannot calculate consistency with fewer than 2 displays\n");
    else {
        for (row=0; row<n_rows; ++row) {
            for (col=0; col<n_cols; ++col) {
                n = cum_mask[row][col];
                if (n > 0) {
                    ++n_occupied;
                    sum = sum + n;
                }
            }
        }
        consist = (sum - n_occupied)/((n_displays-1)*n_occupied);
        consist = consist*'100;
    }
    return(consist);
}

print_specs()
{
    printf("This analysis is based on the following parameters:\n\n");
    printf("Characteristics of display device:\n");
    printf(" Maximum number of characters per row= %d\n\n.n_cols);:
    printf(" Maximum number of rows= %d\n\n.n_rows);:
    printf(" Distances between character centers:\n");
    printf(" Horizontally= %f\n\n.char_width);:
    printf(" Vertically = %f\n\n.char_height);:
    printf("Viewing distance= %f\n\n.visual_angle/TANGENT);:
    printf("Coefficient used in determining groups = %f\n\n.mean_coeff);:
printf("Delimiter between labels and data on display= \%c\n\n",delimiter);