INFORMATION TO USERS

This was produced from a copy of a document sent to us for microfilming. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help you understand 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 you of complete continuity.

2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed the photographer has followed a definite method in "sectioning" the material. 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 any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases we have filmed the best available copy.
FACTORS AFFECTING ITEM ACQUISITION PERFORMANCE IN HIERARCHICAL SYSTEMS: DEPTH VS. BREADTH

The Ohio State University

Copyright 1980 by Miller, Dwight Peter
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 \( \checkmark \).

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 _________________________________________________________

University
Microfilms
International
FACTORS AFFECTING ITEM ACQUISITION PERFORMANCE 
IN HIERARCHICAL SYSTEMS: DEPTH VS. BREADTH 

DISSERTATION 

Presented in Partial Fulfillment of the Requirements for 
the Degree Doctor of Philosophy in the Graduate 
School of The Ohio State University 

By 

Dwight Peter Miller, B.S.E.E., B.A., M.A. 

* * * * * 

The Ohio State University 

1980 

Reading Committee: 

Harvey G. Shulman, Ph.D. 

Richard J. Jagacinski, Ph.D. 

Richard D. Gilson, Ph.D. 

Keith A. Hall, Ph.D. 

Approved By 

Harvey G. Shulman 

Adviser 

Department of Psychology
To my parents Edith and Henry
ACKNOWLEDGMENTS

I wish to express my gratitude to Dr. Harvey G. Shulman for his guidance and encouragement in the conceptual development of the original idea into a practicable plan of research. I would also like to express thanks to Dwayne A. Ball for valuable advice concerning experimental design and statistical analyses throughout the study. Thanks are also owed to Stuart S. Bedoll whose computer programming skills kept the project on schedule and to Diane B. Roth for her patient assistance with the manuscript.
VITA

February 8, 1951 . . . . . . Born - Greenwich Village, New York

1974 . . . . . . . . . . . . B.S.E.E., B.A., Lafayette College, Easton, Pennsylvania

1974-1978 . . . . . . Research Associate, Human Performance Center and Aviation Psychology Laboratory, The Ohio State University, Columbus, Ohio

1977 . . . . . . . . . . . M.A., The Ohio State University, Columbus, Ohio

1978-1979 . . . . . . Associate Research Scientist, Honeywell Man-Machine Sciences Group, Minneapolis, Minnesota

1979-1980 . . . . . . Research Associate, Industrial and Systems Engineering Department, The Ohio State University, Columbus, Ohio

PUBLICATIONS AND PAPERS

A comparison of kinesthetic-tactual and visual displays via a critical tracking task. Human Factors, 1979, 21, #1, 79-86. (with R. J. Jagacinski and R. D. Gilson)


VITA (continued)


FIELDS OF STUDY

Major Field: Experimental Psychology


Minor Field: Human Factors Engineering

Studies in Industrial Safety. Professors Thomas H. Rockwell and George L. Smith


# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>VITA</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xii</td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Review of Relevant Studies</td>
<td>8</td>
</tr>
<tr>
<td>II. APPROACH</td>
<td>14</td>
</tr>
<tr>
<td>Experimental Task</td>
<td>14</td>
</tr>
<tr>
<td>Hierarchical Structure</td>
<td>15</td>
</tr>
<tr>
<td>Choices per Menu (Breadth)</td>
<td>16</td>
</tr>
<tr>
<td>Levels of Menus (Depth)</td>
<td>18</td>
</tr>
<tr>
<td>Item-Category Relationship</td>
<td>19</td>
</tr>
<tr>
<td>III. PERFORMANCE MODEL</td>
<td>22</td>
</tr>
<tr>
<td>Predictive Example</td>
<td>24</td>
</tr>
<tr>
<td>IV. METHOD</td>
<td>28</td>
</tr>
<tr>
<td>Subjects</td>
<td>28</td>
</tr>
<tr>
<td>Task</td>
<td>28</td>
</tr>
<tr>
<td>Design</td>
<td>28</td>
</tr>
<tr>
<td>Materials and Apparatus</td>
<td>30</td>
</tr>
<tr>
<td>Procedure</td>
<td>34</td>
</tr>
<tr>
<td>V. RESULTS</td>
<td>43</td>
</tr>
<tr>
<td>Pretest</td>
<td>43</td>
</tr>
<tr>
<td>Goal Acquisition</td>
<td>43</td>
</tr>
<tr>
<td>Control Tasks</td>
<td>54</td>
</tr>
<tr>
<td>Model Prediction of Goal Acquisition</td>
<td>63</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS (continued)

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>VI. DISCUSSION</td>
<td>73</td>
</tr>
<tr>
<td>Pretest</td>
<td>73</td>
</tr>
<tr>
<td>Goal Acquisition</td>
<td>74</td>
</tr>
<tr>
<td>Control Tasks</td>
<td>81</td>
</tr>
<tr>
<td>Model Prediction of Goal Acquisition Performance</td>
<td>86</td>
</tr>
<tr>
<td>Conclusions</td>
<td>94</td>
</tr>
<tr>
<td>Future Research</td>
<td>96</td>
</tr>
</tbody>
</table>

APPENDIXES

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Word Hierarchies</td>
<td>97</td>
</tr>
<tr>
<td>B. Instructions to Subjects</td>
<td>102</td>
</tr>
</tbody>
</table>

LIST OF REFERENCES             | 120  |
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pretest AT Statistics for Seven Groups</td>
<td>44</td>
</tr>
<tr>
<td>2</td>
<td>Regression Parameters for CM(C) and IM(C)</td>
<td>66</td>
</tr>
<tr>
<td>3</td>
<td>Comparisons of Mean AT with Predicted AT</td>
<td>68</td>
</tr>
<tr>
<td>4</td>
<td>Comparison of Alternative Models Predicting AT</td>
<td>91</td>
</tr>
<tr>
<td>5</td>
<td>Goal Acquisition Error Predictions Based on Control Errors</td>
<td>95</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Prediction of $\hat{A}$ by the Additive Model Using $\hat{C}(C)$ and $\hat{I}(C)$</td>
<td>27</td>
</tr>
<tr>
<td>2</td>
<td>Experimental Design</td>
<td>31</td>
</tr>
<tr>
<td>3</td>
<td>Response Panel Configurations</td>
<td>33</td>
</tr>
<tr>
<td>4</td>
<td>Pretest Word Hierarchy</td>
<td>36</td>
</tr>
<tr>
<td>5</td>
<td>Flowchart of Goal Acquisition Trial Timing Sequence (Example Shown for E4$^3$ Condition)</td>
<td>39</td>
</tr>
<tr>
<td>6</td>
<td>Flowchart of Motor Control Trial Timing Sequence (Example Shown for E4$^3$ and C4$^3$ Conditions)</td>
<td>40</td>
</tr>
<tr>
<td>7</td>
<td>Median Acquisition Times for Six Subjects over Practice in the E2$^6$ Condition</td>
<td>45</td>
</tr>
<tr>
<td>8</td>
<td>Median Acquisition Times for Six Subjects over Practice in the E4$^3$ Condition</td>
<td>46</td>
</tr>
<tr>
<td>9</td>
<td>Median Acquisition Times for Six Subjects over Practice in the E8$^2$ Condition</td>
<td>47</td>
</tr>
<tr>
<td>10</td>
<td>Median Acquisition Times for Six Subjects over Practice in the 64$^1$ Condition</td>
<td>48</td>
</tr>
<tr>
<td>11</td>
<td>Mean Acquisition Times for Each Experimental Group</td>
<td>50</td>
</tr>
<tr>
<td>12</td>
<td>Mean Acquisition Times for Groups Plotted Against Display Choices for Four Levels of Practice</td>
<td>52</td>
</tr>
<tr>
<td>13</td>
<td>Mean Percent Errors for Each Experimental Group</td>
<td>53</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>14</td>
<td>Overall Mean AT and Percent Errors for Experimental Groups</td>
<td>55</td>
</tr>
<tr>
<td>15</td>
<td>Mean Category Match Times for Each Control Group</td>
<td>56</td>
</tr>
<tr>
<td>16</td>
<td>Mean Identity Match Times for Each Control Group</td>
<td>57</td>
</tr>
<tr>
<td>17</td>
<td>Mean Percent Errors for CM and IM Tasks in Control Conditions</td>
<td>59</td>
</tr>
<tr>
<td>18</td>
<td>Overall Mean CM, IM and Percent Errors for Control Groups</td>
<td>61</td>
</tr>
<tr>
<td>19</td>
<td>Mean Motor Control Times for All Seven Conditions</td>
<td>62</td>
</tr>
<tr>
<td>20</td>
<td>Overall Mean Acquisition Times (OMAT) and Original Predicted Acquisition Times (OPAT) vs. Display Choices</td>
<td>64</td>
</tr>
<tr>
<td>21</td>
<td>Observed AT and Predicted AT with CM and IM Regression Lines: Block A</td>
<td>69</td>
</tr>
<tr>
<td>22</td>
<td>Observed AT and Predicted AT with CM and IM Regression Lines: Block B</td>
<td>70</td>
</tr>
<tr>
<td>23</td>
<td>Observed AT and Predicted AT with CM and IM Regression Lines: Block C</td>
<td>71</td>
</tr>
<tr>
<td>24</td>
<td>Observed AT and Predicted AT with CM and IM Regression Lines: Block D</td>
<td>72</td>
</tr>
<tr>
<td>25</td>
<td>Acquisition Time, Percent Errors and Their Sum vs. Display Choices and Levels</td>
<td>79</td>
</tr>
<tr>
<td>26</td>
<td>Motor Control Group Means as a Function of Stimulus Uncertainty</td>
<td>85</td>
</tr>
<tr>
<td>27</td>
<td>Category Match (CM), Identity Match (IM), and Motor Control (MC) Group Means as a Function of Display Choices</td>
<td>89</td>
</tr>
</tbody>
</table>
### List of Figures (continued)

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>Various Model Predictions of Acquisition Time</td>
<td>93</td>
</tr>
<tr>
<td>29</td>
<td>Word Hierarchy for $E8^2$ and $C8^2$ Conditions</td>
<td>98</td>
</tr>
<tr>
<td>30</td>
<td>Word Hierarchy for $E4^3$ and $C4^3$ Conditions</td>
<td>99</td>
</tr>
<tr>
<td>31</td>
<td>Word Hierarchy for $E2^6$ and $C2^6$ Conditions</td>
<td>100</td>
</tr>
</tbody>
</table>
**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRT</td>
<td>Cathode ray tube</td>
</tr>
<tr>
<td>MSC</td>
<td>Multifunction switching control</td>
</tr>
<tr>
<td>RT</td>
<td>Reaction time, response time</td>
</tr>
<tr>
<td>CM</td>
<td>Category match</td>
</tr>
<tr>
<td>IM</td>
<td>Identity match</td>
</tr>
<tr>
<td>AT</td>
<td>Acquisition time</td>
</tr>
<tr>
<td>C</td>
<td>Number of choices on one menu</td>
</tr>
<tr>
<td>L</td>
<td>Number of levels in menu hierarchy</td>
</tr>
<tr>
<td>S</td>
<td>Subject</td>
</tr>
<tr>
<td>AT</td>
<td>Predicted acquisition time</td>
</tr>
<tr>
<td>CM</td>
<td>Predicted category match time</td>
</tr>
<tr>
<td>IM</td>
<td>Predicted identity match time</td>
</tr>
<tr>
<td>E</td>
<td>Experimenter</td>
</tr>
<tr>
<td>E26, E43</td>
<td>Experimental condition performing goal acquisition trials using the 2^6, 4^3, 8^2 or 64^1 hierarchy</td>
</tr>
<tr>
<td>E82, 641</td>
<td>Control condition performing CM and IM tasks using the 2^6, 4^3, or 8^2 hierarchy</td>
</tr>
<tr>
<td>PRT</td>
<td>Partial response time</td>
</tr>
<tr>
<td>MC</td>
<td>Motor control</td>
</tr>
</tbody>
</table>

xii
INTRODUCTION

In an age when digital computer technology not only improves existing systems through various forms of automation but provides for creation of innovative system concepts, the designer of the person-system interface faces unique challenges. Although computerized systems relieve the burden of lower level mundane tasks, the human operator must supervise a greater and more diverse set of higher-level operations, make tactical decisions and monitor larger portions of the system for failures. Thus, in some cases where expansion of system capability is accompanied by cutbacks in personnel the net effect is an increase in operator workload. So now, more than ever, the person-system interface must provide for quick and accurate information acquisition, function selection, and control actuation to maintain system efficiency.

Prior to the early 1970's several common techniques were used to handle proliferation of system displays and controls. Spatial arrangements were redesigned such that the most important sources were in the direct line of sight, displays were reconfigured to be more compatible with the associated control responses, and sense modalities other than vision were used for presentation of information. This compounded "horizontal" distribution of information sources is exemplified by early versions of today's advanced jet fighter aircraft cockpits. Until only recently, expansion of functional capability brought with it proliferation of corresponding dedicated display and control hardware to the already cluttered cockpit instrument panel. Broadening the array of
information sources ensured equal accessibility but required the operator
to sort through numerous sources to retrieve the desired information.
The presence of all displays and controls did not help to reduce the
danger of interpreting or actuating the wrong device, but encouraged it.
The horizontal configuration's greatest asset is that any display or
control is immediately accessible, limited only by the operator's
familiarity with the format of the panel (Nevins and Johnson, 1972).

An alternative strategy is to distribute information such that the
operator has only the currently relevant portion of the system with which
to interact at any given time. Unused information can be stowed out
of sight within a more "vertical" structure and called up when needed
by the operator (Nevins and Johnson, 1972). Replacing dedicated dis-
plays and control devices with generalized multipurpose devices allows
many subsystems to timeshare common input and output hardware. This
method helped solve the fighter cockpit situation when panel designers
literally ran out of physical space. Using cathode ray tube (CRT) dis-
plays and computer generated graphics, cockpit designers replaced a
multitude of dedicated displays with a few CRTS. Essentially an elec-
tronic blackboard, the CRT can display information in an endless
variety of two dimensional formats, providing a flexible multipurpose
display medium.

The analogous development in control hardware is the multifunction
switching control (MSC) which can replace numerous panels of discrete
dedicated switches with a common array of multipurpose switches shared
for all system inputs and commands. Each switch performs a variety of
switching functions at different times by having the flight computer
differentially interpret its inputs and by announcing its current
function through a variable format label. The computer restricts the control options available to the operator by displaying only those controls which are relevant to his current task.

Hence, the fighter pilot's task of acquiring information has evolved from the act of changing visual fixation to actively telling the computerized system which information source he wishes to see. Certain safety considerations require dedicated displays for immediate access and redundancy such that most practical solutions include a combination of vertical and horizontal information structures.

Accessibility of various displays and controls in a mixed horizontal and vertical system is dictated by control logic which determines the path by which the operator can arrive at the displays he wishes to see and the control options he wishes to have at his disposal. A simple example may best illustrate the notion of control logic. Older model televisions with rotary channel selectors allow complete parallel (horizontal) access to all controls except channel selection which requires sequential changes in channel. Thus, in going from channel 2 to channel 5, one must pass through channels 3 and 4. This represents vertical structure in that channel 5 cannot be accessed until 3 and 4 are. Newer sets with pushbutton tuners allow independent parallel access so that channels can be selected in any sequence. These control logics are the result of particular hardware chosen for the television design. In more complex systems the control logic is a result of considering inherent system organization, user requirements and relative priorities of the various functions. When numerous functions or modes are accessed through a limited channel device such as a MSC, a hierarchy of several levels of indenture organizes the control logic by grouping
related functions on different branches of a tree structure.

When the Air Force Avionics Laboratory attempted to fully integrate the avionics systems for single-pilot advanced fighter aircraft, the resulting Digital Avionics Information System (DAIS) organized the entire control logic into four levels of hierarchical indenture. The first level comprised dedicated controls for the major areas pilots have to manage such as communications, navigation, sensors, aircraft, and weapons. The second and third levels incorporated 196 active DAIS functions into software for presentation on two 16-button multifunction keyboards. The fourth level was reserved for digital entry on a dedicated touchpad. In order to change the communications radio frequency the pilot would go through the following sequence of inputs: First he would press the communications button on the dedicated mode-select panel. The system would then acknowledge his input by backlighting the chosen button, display a page of communications status on one of the multipurpose CRT displays, and present a set of communications options on the multifunction keyboard. The pilot then would choose which radio he wanted by pushing the button reading "UHF." The keyboard legends would change once more to provide him with action options such as frequency and channel selection. After selecting "FREQ SEL" he would enter the frequency digits using the dedicated touchpad (Armstrong, 1975). This sequence is considerably more involved than reaching for the frequency selection knobs on the number one communications radio and dialing to a new frequency. Similar multi-input sequences are necessary to accomplish other tasks on the DAIS.
If MSCs are well designed, they can minimize the number of switching operations required and guide the operator through complicated switching operations necessary for control of sophisticated systems. Graham (1974) developed a step-by-step procedure for the design of MSCs intended to optimize the tradeoff between number of switching controls required and the associated operator workload. He took into consideration functional flow of the operators' tasks, display priorities and control function urgency, frequency of use and criticality in assigning functions to levels of indenture. In a similar effort, Calhoun (1978) empirically determined control logic design criteria for MSCs through an iterative process of redesigning the cockpit of an advanced attack aircraft simulator and soliciting pilot evaluation. The following criteria and guidelines were among those developed:

1) Highly critical functions should be eliminated from the multifunction control.

2) Frequently used and emergency functions should have dedicated controls.

3) Switch legends should be specific and unambiguous.

4) Control logic should be tailored such that options likely to be used during a particular flight phase should appear together.

5) Related functions should be assigned to adjacent switches.

6) Hand motion should be minimized in sequential operations.

7) Functions should be consistently assigned to the same switch and potential sequential operations should not be restricted.

8) Unnecessary digit entries should be eliminated and selections should be verified with feedback.

9) Provisions should be made for return to previous menus when entry
errors are made.

Graham (1974) and Calhoun (1978) offer valuable suggestions for procedures and criteria to use in designing an MSC.

One consideration is lacking in both of these analyses. Neither author discusses how many levels of indenture should be used nor how many options should be listed at each selection point in the structure. Calhoun states that "no function or piece of data should be more than four switch hits removed from the first menu or display," but doesn't provide any basis for this guideline. It seems reasonable to suggest some limit for levels of indenture so that operators don't spend an inordinate amount of time punching buttons seeking access to goal functions. However, with a given finite number of functions to be accessed, the curtailment of levels of indenture means more functions must share each level. This could cause some inconsistencies with Calhoun's other guidelines as functions which logically should reside in subordinate levels are folded up into higher levels of the control logic. Presenting the operator with a multitude of options, some of which may be categorically different may have its own deleterious effects upon selection performance. The tradeoff between the number of levels of indenture and the numerosity of options at any decision node should be systematically investigated. Most certainly either extreme represents an implausible solution, but perhaps an optimal balance between depth and breadth of the control logic will enhance system comprehension and function acquisition performance. Perhaps the determinant factor is not balance between depth and breadth but the absolute practical limit on either one of these dimensions.
This consideration applies not only to MSCs but to almost every complex system capable of numerous processes or functions. Modern power plant control rooms are changing from large rooms with thousands of displays and controls crammed into control panels to several multi-purpose CRTs capable of displaying any system indication called by the operator or "forced" to the display by a particular critical event. In order to "call up" specific elements of the complex system, the controlling keyboard must either have dedicated keys corresponding to each element or a means of concatenating inputs on a smaller set of more generalizable keys. Likewise most interactive computer terminals employ CRTs and typewriter-style keyboards. Except for a few dedicated keys which are used frequently and have only one purpose, these keyboards allow an infinite variety of inputs constructed as strings of alphanumeric characters. These strings can be used to call up data, insert data or command a specific function to be performed. If the system is very complex the number of valid input strings can be quite large. Operators must either commit them to memory or look them up in a catalogue.

A more modern approach is for the system to present a "menu" of items from which the user chooses. Just as with the MSC a few sequential menu selections can specify one function from the entire set of possibilities. Here, the dimension of depth translates to the number of levels of menus and that of breadth to number of selections on each menu. The most recent CRT terminal designs often do away with keyboards altogether. A touch panel superimposed over the face of the CRT transforms that surface into a variable-format keyboard. The operator simply "touches" the item he wants with his finger and the touch panel senses
the location of his contact with the screen. The computer then correlates the input location and the corresponding menu item through software control. The touch display was originally conceived at the Royal Radar Establishment in an attempt to improve man-machine communications in an air traffic control data processing system (Johnson, 1967). In the past decade, touch displays have experienced diverse applications ranging from industrial process control to home computers. Touch displays represent the ultimate in MSCs and are potentially the height of man-computer sybiosis short of speech recognition and synthesis communication. Since menu selection is the basis for control, intelligent design of menus and their control logic is paramount in optimizing this man-machine interface. It is the intention of this research effort to contribute to the development of human factors guidelines for this design activity.

Review of Relevant Studies

One responsibility of scientific research is to provide the context or setting for the current work and to justify the effort. This is done by reviewing and aggregating previous related studies, explicating the relationship with past endeavors and suggesting future approaches to succeed the current work. In providing the context for the present research, five key relevant papers will be briefly discussed.

Nevins and Johnson (1972) explored display and control techniques for complex space-bourne interactive computers. In the first application of its kind, the authors attempted to employ a mixed (vertical and horizontal) control logic in designing a prototypical control/display system. Data regarding systems status were structured into four functional units to maximize interpretability and availability: (1) major
mission phase sequences, (2) vehicle status, (3) subsystems status and (4) special computation routines. A network tree-like structure with links across branches enabled access to special data during major mission sequences. A touch display was implemented for ease of input and no piece of data was more than four key presses from the top level menu or "Prime" display. Several control functions were shown as dedicated "keys" on most displays to aid the user in navigating through the control logic. A "Recycle to Prime" key allowed return to the top level menu. A "Return to Last Display" key allowed quick correction for miskeys by the operator. A "Call Data" key provided a menu of subsystems and functional units under which a library of data was stored in memory.

A particularly nice human factors feature was included. When the "Call Data" matrix was called for, the prior menu (from which it was selected) was scaled down and presented in the upper left corner of the CRT to aid the operator in keeping track of his location in the sequence. Among the advantages over conventional designs, the prototype interactive control scheme offered: no attentional shifts between information displayed and choice execution, fewer entries necessary to retrieve data, no numeric codes needed to identify input commands and inclusion of data which would normally require a large number of dedicated displays.

A similar type of interactive cockpit touch input/output system was simulated and evaluated by Gärtnert and Holzhausen (1977). Menu selection hierarchies and virtual keyboards were used to control an integrated flight management system similar to the Litton LTN 72-R used by German Lufthansa airlines.

Among the ergonomic problems identified was the lack of tactile feedback when using the virtual keyboard on the CRT surface.
Experimentation proved that larger switch diameters decreased activation time and the authors inferred that virtual switches should employ larger sizes than their electromechanical counterparts to aid visual feedback of key actuation. Görtner and Holzhausen also mentioned that access paths should incorporate an "ergonomically designed menu technique" to minimize the delay associated with sequencing through the hierarchy of levels to acquire a specific function. A "random and sequential" menu page selection hierarchy was used to demonstrate the possibility of overloading the pilot with manual data input procedures. Basically a three-level hierarchy with random access to the second and third levels, the third level was made up of "chapters" comprised of pages that only could be accessed sequentially. This organization is typical of systems offering parcels of substantial data grouped under singular descriptor terms as in computerized business accounts or literature search data bases.

In a more empirical vein, Barnard, Morton, Long and Ottley (1977) studied menu organization techniques and demonstrated that the way information is presented on a menu affects acquisition time for locating a particular goal item. The task was to read a goal item from a randomized list, find the item on an organized menu and record an arbitrary number paired with the goal item. Geographic organization led to faster performance than alphabetic organization when European countries served as target items and the subject sample was knowledgeable on geography. Similarly, semantically categorized menus of household items were superior to alphabetized listings for housewives performing an identical task. Although a slight detriment in performance was found in progressing from one form of organization to another, the authors
concluded that "appropriate use of an individual's knowledge of a subject matter, when incorporated in the layout of a display, can greatly facilitate his finding of information upon it."

Although this study dealt mainly with single display search performance, the benefit of semantically organizing information can clearly be seen. Subjects searched the display for the appropriate category heading, and then searched the underlying list for the goal item. This task is identical to stepping through a two-level hierarchy except no manual input was needed to call up the selected list. Intelligent use of solid category-item relationships should enhance goal acquisition in multi-level hierarchies as well.

Another display organization study compared batch and sequential presentation of response alternatives on a CRT display. Baker and Goldstein (1966) had subjects learn a verbal "maze" constructed from a hierarchy of nonsense syllables with completely arbitrary relationships. The task involved searching through the maze for a goal object by choosing among the nonsense syllables offered at each branching of the hierarchy. A light pen was used to select the desired trigram from the CRT display. After 16 practice trials of random exploration of the maze each subject was informed of eight goal trigrams and given 16 practice trials to find them. Since the goal items resided at various levels of the maze, correct recovery sequences varied in number of inputs. Following practice, subjects were required to retrieve specific goal objects (displayed at the top of the CRT) continuously until all eight could be retrieved without error.

In the batch display condition, the goal trigram was presented on the top of the CRT. The available choice alternatives for the particular
decision node were displayed underneath the goal, and the bottom half displayed a 4 x 5 matrix of all the maze elements from which the subject identified his choice with the light pen. To prevent utilization of spatial cues the positions of the stimuli were randomized within their respective regions after each trial. The sequential display condition showed the goal object and the available choice alternatives for the particular decision node, from which the subject chose using the light pen. Hence the subject selected directly from a maximum of four valid alternatives on the sequential display as opposed to a matrix of 20 items which included the valid set on the batch display.

Although no differences were found in number of trials to criterion, subjects using the sequential display reached criterion performance in significantly less time than those using the batch display. Furthermore, subjects demonstrated reliably faster goal acquisition using the sequential display during the criterion trials. The authors concluded that the increased density of displayed information led to increased display search times.

This study suggests to the human factors designer that he should omit irrelevant information from a menu but does not address the issue of how much relevant information is practicable. One has to wonder also whether similar effects would hold for meaningful stimuli and semantic relationships among the stimuli within a verbal structure.

Kammann (1972) compared four instructional techniques for teaching employees 16 different telephone dialing tasks. Two were traditional paragraph types, one which had been previously used for eight years and the other a rewritten version with improved wording. A new type consisted of a hierarchy of binary questions in a flowchart of boxes and
arrows with a maximum path length of seven decisions to arrive at the most remote instruction set (F1). A second flowchart (F2) had three and four-choice branches and a maximum path length of three decisions. The latter two instruction types relate to the present study. Both proved to be equally superior to the conventional instructions in terms of error rates for "high-level" employees at Bell Laboratories. For housewives, F2 was superior to F1 and both were superior to conventional instructions. The author suggests that shortening total information and decisional steps in a path is more valuable than a fully explicit hierarchy of binary decisions. Hence we have one source in the literature demonstrating an instance where a structure with reasonable breadth and depth is easier to use than one with little breadth and great depth for a particular sample of users.

Taken together these studies demonstrate the need for the current research. The first two demonstrate the benefits that can be derived from integrated computerized touch input/output systems when applied to high operator workload situations. The next two studies discussed show how organizing a page of information can improve item acquisition performance. Finally, Kammann's work suggests that manipulation of the hierarchical structure of a complex system can influence item acquisition performance. The present study will assess performance tradeoffs for system depth versus breadth. Consideration is given to psychological principles affecting performance in the goal acquisition task. Those thought to be especially salient are discussed next in attempt to construct reasonable hypotheses.
APPROACH

It is desirable to reflect on the cognitive psychological literature to answer the question of depth versus breadth. However, it is not a simple matter to identify the salient, relevant influences, estimate their individual effects and then calculate the interactive effect of their sum upon goal acquisition performance. Cognitive ergonomics is sufficient for simple problems, but the goal acquisition task is too complex for straightforward human factors inferences and armchair organization. The goal acquisition task should be abstracted from the "real world" task, analyzed by its component processes and empirically investigated in a laboratory setting.

Experimental Task

The experimental task chosen for the current experiment simulates goal acquisition on a multi-level hierarchical system. Verbal hierarchies of various configurations were constructed for a set of goal items residing at the lowest level of a set of semantically nested terms. Goal items (common English words) were presented to the subject on a CRT display in a random sequence. After each goal item is presented and the first level menu is displayed, the subject chose the appropriate category descriptor word by pushing a button on a response panel adjacent to the word's location. The appropriate second-level menu then appeared and the sequence continued until the goal item was acquired. Partial and total response times were collected for every trial.
Emphasis was placed on speeded performance with minimal errors. A discussion of salient features of the task and their effects upon task performance follows:

Hierarchical Structure

The first issue to be discussed is the organizational compatibility of system structure and acquisition procedures with inherent human cognitive structures and processes of information organization and retrieval. Durding, Becker and Gould (1977) investigated how people organize information by having subjects organize word sets with minimal instructions. The authors found that subjects "had available" all of the organizational structures tested: hierarchies, networks, lists and tables. In addition, the structures used reflected those inherent in the data that were organized. Durding et al. suggested that a database in any system should conform to the semantic relationships found in the data elements and that if data conforms to a hierarchical structure the user will be capable of mental manipulations of the data.

Several authors support the notion of hierarchical internal structures in the more traditional psychology literature. Bower, Clark, Lesgold and Winzenz (1969) found that a hierarchical structure given during learning of words aids recall and recognition substantially over randomly organized lists of the same words. Collins and Quillian (1969) analyzed true and false responses to simple questions and supported the proposition that long-term memory is economically organized in a hierarchical fashion. Broadbent, Cooper and Broadbent (1978) found equivalent recall for common nouns presented in matrices and hierarchies and concluded that recall benefit is due to grouping of items rather than use of category cue words. These studies support the idea that
people can deal with hierarchically organized information proficiently and that it is compatible with extant long-term memorial structures.

**Choices Per Menu (Breadth)**

Considering the task of finding a target word among several words in a displayed list, one turns to literature in the field of visual search. Visual search research typically involves searching a display containing several items for one or several target items. Unfortunately, the stimuli typically used are geometric shapes or capital letters rather than words. Words are the stimuli of verbal learning researchers who usually employ tasks involving several memory set items and one or two display items. However, a few studies which span this apparent gap can be consulted to assess the effect of display size upon goal acquisition.

Neisser and Beller (1965) had subjects search through lists of 50 words for a specific target item (the word "Monday") and for a member of a target set (an animal). Over the first eight days of practice, subjects could scan the lists at approximately .130 seconds per word for specific words and .250 seconds per word for target set items. Considering error rates of 10 percent and 22 percent respectively, these rates may be slightly faster than when error rates are confined to smaller levels. Shulman (1971) had subjects search three columns of 16 words each for members of target categories "living things" and "geographical locations." Scan rates averaged 1.10 seconds per word but half of the words were targets and subjects had to circle a "yes" or "no" response on the paper adjacent to the word. The unconventional response mode probably inflated the scan rate estimates considerably.
Since the present experimental task involves presentation of word lists (menus) on a CRT, studies involving presentations of fewer items for shorter duration were also consulted. Estes and Wessel (1966) evaluated two-alternative forced choice responses to tachistoscopically presented fields of capital letters. Every display included either target "B" or "F". Correct response times increased linearly with the number of displayed letters. When corrected for differing display durations and plotted, response time increased from .59 seconds at one letter to .90 seconds at 16 letters, producing a slope of .022 seconds per letter, and an intercept of .56 seconds. Similarly, Nickerson (1966) used 1, 2 or 4 letters as stimuli with one memorized target item in a binary classification task. Response times for positive responses again increased linearly with display size giving a slope of .063 seconds per letter and a .40 seconds intercept. Atkinson, Holmgren and Juola (1969) confirmed these data for letter stimuli but got a consistent slope of about .025 seconds per letter across all levels of practice. Briggs and Blaha (1969) used eight-sided random figures as stimuli in a similar binary classification paradigm and obtained similar results. After three days of practice performance virtually stabilized to a linear function of number of displayed items with a slope of .037 seconds per figure and an intercept of .45 seconds.

The data reviewed above demonstrate clearly that positive response times increase proportionally and linearly with display size. The slope of these functions vary with the type of stimuli used and method of presentation. However, regardless of stimuli used and the process model advocated by the author, in each case a linear relationship obtains. This effect translates directly to the present experimental task of finding a
target word in a displayed menu. As the menu size increases, so does response time for selecting the correct item.

Levels of Menus (Depth)

The goal acquisition task requires interaction with several menus in succession. One can only speculate on the possible difficulties encountered with this serial process. Minimally, the operator must remember the goal item in order to choose the appropriate alternatives from each successive menu. As the user becomes more familiar with the hierarchical system, he may begin to learn paths to goals and the overall structure, not unlike learning a maze or constructing a cognitive map. If cognitive maps are developed, the user should have no difficulty remembering paths to goal items as the spatial aspects of the system will aid traveling through the hierarchy. However, if paths are remembered as strings of menu items, limitations of working memory may hinder performance.

The phenomenon of a working memory buffer of limited capacity or short-term memory (STM) is well documented in the literature (Miller, 1956; Atkinson and Shiffrin, 1968). This inability to keep track of more than a few items may influence the ability of a user to step through a multi-level hierarchical system. In order to know where one is, one must remember how one arrived. In other words, remembering the sequence of steps in the path leading to the current location may be critical to remaining cognizant of current location, the next immediate step and the ultimate goal location. Short-term memory limitations suggest that path lengths greater than about five items will be difficult to remember. This has obvious implications for limiting the number of levels or depth of a hierarchical control logic.
A much more obvious result of increased levels of menus is the need for several overt responses. Each time an item is chosen from a menu, the user must communicate his choice to the system by pushing a button, or by some other means. Each response takes a finite time to code and execute. Upon completion of the response, reorientation to the subsequent menu takes place. Obviously, the more responses required to acquire a goal item, the more time will be taken in responding to the system and acquiring the goal item.

**Item-Category Relationship**

Except for the lowest hierarchical level, when a menu is searched for the appropriate selection, the user has to match the goal item with a category descriptor. For instance, if the goal word is ROSE, an appropriate menu selection may be PLANT at one level, and FLOWER at a lower level. Ultimately, the word ROSE will appear on the final menu. Variables such as category size and semantic relationships affect verbal performance and provide implications for the design of word hierarchies. Landauer and Freedman (1968) found that reaction time (RT) increased with the size of categories in tasks where item-category relationships were tested as valid or invalid. Collins and Quillian (1969) explained this effect as caused by semantic distance. When semantically proximal stimuli were compared, responses were faster than for semantically distant items.

In attempting to clarify a theory of semantic memory, Rips, Shoben and Smith (1973) had subjects verify semantic relations in simple statements such as "a ROBIN is a BIRD." They found that verification was faster when the target category was a direct superordinate (BIRD) than a higher level superordinate (ANIMAL). Although the effect was
relatively strong (a 10 percent increase in response time) it was not reliable across all category descriptors used.

Regarding category cue words, Broadbent and Broadbent (1978) found personal descriptors superior to those provided by others when classifying items for subsequent retrieval. This suggests that at the very least, category cue words for a system should be reasonable descriptors to which all users can relate.

The potential effects of different item-category relationships on goal acquisition are uncertain. The semantic distance effect may tend to make higher level menu selections more difficult than lower level selections. Hence, the partial response times for an acquisition trial may decrease as the user approaches the lowest hierarchical level.

One consistent finding is that searching for a particular word is a faster process than searching for a member of a category. Neisser and Beller (1965) found that people found the word "Monday" in a list faster than they found a member of the category "Animal." In fact, the scan rate for categories was .25 seconds per word while that for specific words was .13 seconds per word. Shulman (1971) got similar results in his search task. Subjects consistently took longer to identify living things than "words including the letter "A". The scan rates per word were 1.10 and 1.0 seconds respectively. This 0.1-0.12 seconds per word difference may translate to the goal acquisition task paradigm. Finding a category descriptor for a goal item could take 0.10-0.12 seconds longer per word scanned than finding the goal item itself in a list. Thus, an additional penalty for a hierarchically deep system is increased processing time for every category match necessary.
The variables already discussed are predicted to have significant effect upon goal acquisition performance in hierarchical systems. They are also under the control of the system designer in that he has great latitude in arranging system elements, their labels and their relationships to each other. In order to assess the relative effects of these parameters, it is desirable to control and neutralize the influence of other variables which may also affect task performance. Therefore variables such as word length, word familiarity, display size and arrangement, letter size, response coding, means of response and practice will be controlled for the extent possible in the present methodology.
The goal acquisition task may be considered a serial task of several identifiable components. Consider a three-level hierarchy having menus with four choices each. Choosing the target on the first menu consists of matching the memorized goal item with the appropriate category descriptor. This process will be called a category match (CM) for convenience. A CM involves comparing each menu item on the display with the goal item in memory until a match is perceived. The subsequent selection of the target category by an overt response calls up the underlying menu which again consists of four categories. The subject makes a second CM, further identifying the goal. The final menu consists of four items, one of which is the goal, assuming all previous CMs were correct. Here, the subject chooses the identical item from the list that matches the memorized goal item. This process will be called an identity match (IM). Regardless of system structure, an IM is always the concluding response in the goal acquisition task. Therefore, assuming minimal interaction among components, goal acquisition time (AT) will be the sum of the component times:

\[ AT = CM_1 + CM_2 + IM \]

More generally speaking, for a hierarchy with L levels, the number of CM components equals one less than the total number of levels:

\[ AT(L) = CM_1 + CM_2 + CM_3 \ldots + CM_{L-1} + IM \]
As has already been demonstrated in previous research, search time increases linearly with the number of stimuli presented. This relationship should hold for both CMs and IMs. As the number of categories displayed increases, the time taken to find the target category should increase proportionately. The same principle holds for finding a target item in the IM task. Using C to denote number of choices in the display, the times for CM and IM can be expressed by the following linear equations:

\[ CM(C) = a + \alpha C \]
\[ IM(C) = b + \beta C \]

The lowercase a and b represent ordinate intercepts while \( \alpha \) and \( \beta \) denote processing rates which are constant over a broad range of C.

Returning to the model dealing with acquisition time (AT), and taking parameter C into account, the model becomes:

\[ \hat{AT}(L,C) = (L-1)CM(C) + IM(C) \]

This model assumes equal C for the entire goal acquisition task. It also assumes that hierarchical distance will have little effect upon CM times. Hierarchical distance is the number of levels separating the target category at a given level and the lowest or identity level. A more general form of the model which does not assume equal C at each level can be written:

\[ \hat{AT}(L,C) = \sum_{n=1}^{L-1} CM_n(C) + IM(C) \] (1)
A desirable feature of the \( \hat{AT} \) model, is that the two straight lines describing CM and IM as functions of \( C \) are sufficient to predict AT for any hierarchical configuration. Furthermore, the relative values of \( \alpha \) and \( \beta \) can predict the optimal values of \( L \) and \( C \) to minimize the objective function \( \hat{AT}(L,C) \). This potential can be best illustrated with a specific predictive example.

**Predictive Example**

Suppose a system contains a moderate number of elements. Sixty-four is a good number for this illustration. These 64 elements are the goal items for the acquisition task. They can be accessed via several different hierarchies of category descriptors. The simplest is a single-level hierarchy, or a list of the 64 elements. Thus, with 64 choices \( C = 64 \) and only one level \( L = 1 \), the hierarchy can be expressed as \( C^L = 64 \) or \( 64^1 = 64 \). A second way of accessing the 64 goal items is to subdivide them into 8 groups of 8 each. Identification of the desired element is made by identifying the appropriate group \( (CM(8)) \), and then picking the desired element from the list of 8 \( (IM(8)) \). In this case \( C=8 \) and \( L=2 \), giving \( C^L = 8^2 = 64 \). As the 64 elements are subdivided into smaller groups, the hierarchy of categories necessary to identify then grows vertically giving \( C^L = 4^3 = 64 \) and \( C^L = 2^6 = 64 \). These four hierarchies demonstrate how breadth can be traded for depth, satisfying the constraints that \( C \) is equal at all levels, and that 64 elements comprise the lowest level of the system. This set of four hierarchies was used in the present experimentation.

According to the AT additive model, two functions are needed to predict goal acquisition performance. These are: CM as a function of \( C \) \( (CM(C)) \), and IM as a function of \( C \) \( (IM(C)) \). From the studies reviewed
estimates of these functions can be made. Using common English words as stimuli the IM function should be on the order of:

\[ \hat{\text{IM}}(C) = .40 + .08(C) \] (2)

Thus the intercept will be .40 seconds and the processing rate will be .08 seconds per word displayed. The CM function will necessarily be steeper as the task involves item-category matches rather than identity matches. The CM function should be on the order of:

\[ \hat{\text{CM}}(C) = .40 + .17(C) \] (3)

The intercept is equivalent to that for IM(C) indicating that it may represent stimulus encoding and response execution phases of the task. The slope of .17 seconds per word is about 0.1 seconds greater than the IM(C) slope reflecting a more complex processing task. Both of these linear functions are plotted in Figure 1.

Using the AT additive model and the predicted values of CM and IM from the two linear functions, AT(L,C) can be estimated for the hierarchies 2^6, 4^3, 8^2 and 64^1.

For example, \( \hat{\text{AT}}(3,4) \) can be predicted for the 4^3 hierarchy using the following calculations:

\[
\begin{align*}
\text{IM}(4) &= .40 + .08(4) \\
\text{IM}(4) &= .72 \text{ seconds} \\
\text{CM}(4) &= .40 + .17(4) \\
\text{CM}(4) &= 1.08 \text{ seconds}
\end{align*}
\]
\[
\hat{AT}(3,4) = \sum_{n=1}^{2} CM_n(4) + IM(4)
\]
\[
\hat{AT}(3,4) = CM_1(4) + CM_2(4) + IM(4)
\]
\[
\hat{AT}(3,4) = 1.08 + 1.08 + .72
\]
\[
\hat{AT}(3,4) = 2.88 \text{ seconds}
\]

Estimates of AT for the other three hierarchical configurations can be calculated using the same mathematical procedure. This creates a U-shaped curve with a minimum somewhere between \(C = 4\) and \(C = 8\) (Figure 1).

Thus, it is obvious that at the extremes of depth (\(2^6\)) and breadth (\(64^1\)), goal acquisition performance drops off and \(\hat{AT}\) increases. In the middle, where more balance is attained, performance is maximized.

The minimum of the actual, observed AT(\(L,C\)) curve will be determined by the ratio of the slopes of the actual CM(C) and IM(C) functions. This prediction and the adequacy of the AT additive model was tested by the experimentation described in the following sections.
Figure 1. Prediction of $\hat{A}_T$ by the additive model using $\hat{C}_M(C)$ and $\hat{I}_M(C)$.
METHOD

Subjects

Twenty-three males and 19 females from the campus community served as subjects (Ss). Thirty-five responded to a newspaper advertisement and were paid $3.00 an hour for their participation. Seven were recruited from the introductory psychology course S pool and received class credit for participating. All Ss competed for a $5.00 bonus on the pretest session as well as a $5.00 bonus for best performance within the assigned condition.

Design

A between-S design evaluated the effects of hierarchical structure and the validity of the additive AT model. Four experimental conditions had Ss doing goal acquisition tasks by interacting with four different hierarchical structures, each having 64 goal items at the lowermost level. Each structure also had a consistent number of choices (C) at each level (L). All structures conformed with the constraint that \( C^L = 64 \). The deepest structure was \( 2^6 \), which represented binary choices at six different levels. The more balanced structures were \( 4^3 \) and \( 8^2 \). The broadest structure was \( 64^1 \), which manifested the extreme in which all goal items were present on one single menu. Given that 64 goal items represent a system of moderate size, the four structures effectively spanned the depth-breadth tradeoff spectrum. The four experimental groups (\( E2^6, E4^3, E8^2, 64^1 \)) each consisted of six Ss.
Experimental Ss performed goal acquisition trials by searching through the verbal hierarchy for a given goal word. The goal word, randomly selected from the lowest level was presented on a CRT, followed by the highest level menu of category descriptor terms. The S's task was to choose the appropriate category as each successive menu level was presented by pressing its corresponding pushbutton mounted on the CRT. At the lowest level, S selected the goal word from the final menu. Each goal acquisition trial ended with performance feedback for the S.

In order to generate data for the two predictive functions \( \hat{g}(C) \) and \( \hat{m}(C) \), three control groups performed CM and IM tasks which were equivalent to the component sub-tasks of the goal acquisition task in the experimental goal acquisition conditions. For each experimental S that performed a goal acquisition task, a control S performed single-frame CM and IM components of the task taken from the same hierarchical structure. The CM and IM components were independent trials appearing randomly within a trial block. Just as their experimental counterparts the control groups C2^6, C4^3 and C8^2 each had six Ss. For the 64^1 structure the experimental and control groups were one and the same. Thus, a total of seven groups of six people totalling 42 Ss comprised the entire between-S design.

All 42 Ss participated in a pretest session before returning for the assigned condition. A simplified two-level hierarchy was used to instruct Ss on the goal acquisition task and to provide baseline data for group assignment. In addition, all 42 Ss performed an adjunct task between blocks of word processing trials. Each responded to a non-verbal stimulus randomly presented in one of the C menu locations on the display.
screen by pushing the corresponding response button. This motor control (MC) task was used to estimate the contributions of stimulus encoding and response execution for the menu selection tasks employed in the goal acquisition and control tasks. All tasks will be explained in detail in the procedure section.

Practice effects were studied by having Ss go through the complete hierarchy (64 goal acquisition trials) four times. Subjects performed two blocks of 64 trials on each of two consecutive days allowing four levels of practice to be studied. Since Ss performed the experimental or control tasks in 32-trial blocks, many analyses were done on 8 32-trial blocks rather than on 4 64-trial blocks (see Figure 2).

**Materials and Apparatus**

Stimuli – Members of a single set of 64 common English words served as search targets in all conditions. These words were related by category names, also common English words superordinate of the goal words in the hierarchical structure. The semantic structure underlying the Dewey Decimal classification system was used as a source for all stimuli, since it provided sufficient breadth and depth for the selection of structure and stimuli. Once the $2^6$ structure was formed, the $4^3$ and $8^2$ hierarchies were developed as subsets, using category descriptors from appropriate levels of the $2^6$ structure. The $64^1$ structure was simply a list of the 64 goal words without category descriptors.

The major constraint placed on the words used was that they form valid semantic hierarchies with valid superset-subset associations. To control for word length and word familiarity, across conditions the goal words (nouns and proper nouns) were identical for each hierarchy. Meeting these two constraints precluded others desirable in visual search
Experimental Control

Hierarchical Structure

32GA, 16MC
Break
32GA, 16MC
n=6

160CM, 32IM, 16MC
Break
160CM, 32IM, 16MC
n=6

32GA, 16MC
Break
32GA, 16MC
n=6

64CM, 32IM, 16MC
Break
64CM, 32IM, 16MC
n=6

32GA, 16MC
Break
32GA, 16MC
n=6

32CM, 32IM, 16MC
Break
32CM, 32IM, 16MC
n=6

32GA, 16MC
Break
32GA, 16MC
n=6

32CH, 32IM, 16MC
Break
32CH, 32IM, 16MC
n=6

GA = Goal Acquisition
MC = Motor Control
CM = Category Match
IM = Identity Match

Figure 2. Experimental design
experimentation. For example, distribution of word length and familiarity across levels of each hierarchy could not be counterbalanced. (See Appendix A for diagrams of the four word hierarchies.)

**Experimental Terminal** - The acquisition and control tasks were controlled by a Data General Nova 3/12 minicomputer interfaced with a Hitachi Model TWU-72 12-inch raster scan CRT display and a pushbutton response panel. The word stimuli were displayed in upper case letters measuring .25 inches (.63 cm) tall by .094 inches (.24 cm) wide, displayed as black-on-white video. In all conditions but 64\(^1\), the stimuli were presented along the right and left edges of the CRT display, separated by 7.5 inches (19 cm) horizontally and 2.0 inches (5.1 cm) vertically. The S sat in a dimly lit cubicle, approximately 20 inches (.51 m) from the CRT display. This arrangement afforded a visual angle of .716 degrees vertical and .269 degrees horizontal for each letter. The shortest word had three letters subtending 1.074 horizontal degrees visual angle, while the longest had 11 letters subtending 3.933 degrees.

The response panel was custom-built for this experiment. It was mounted directly on the face of the CRT display with a 9.0 x 7.0 inch (23.8 x 17.8 cm) hole cut for video viewing. Response buttons were mounted .31 inches (.79 cm) from the edge of the rectangular viewing hole in slots allowing vertical positioning next to the stimulus fields. The buttons were Robert Shaw \# 701-1052-00 momentary pushbuttons, which provided good tactile feedback and an audible "click" when depressed. The terminal rested on a table 30 inches (.76 m) above the floor and was tilted 10 degrees back from vertical for perpendicular viewing.

Four response panel configurations were necessary for the seven groups of Ss (See Figure 3). Due to physical limitations of the
Figure 3. Response panel configurations
apparatus, Ss in the 64\(^1\) conditions selected the block of eight words containing the goal word rather than identifying the actual goal word. This was done via pushbuttons mounted along the top and bottom of the video field. Pushbuttons not used (C < 8) were recessed behind the front surface of the response panel and covered with plastic tape matching the color of the panel. Likewise, positioning slots not occupied by pushbuttons were covered with tape. Thus each S interacted with response panel containing only the C response buttons aligned with C stimulus fields on the CRT.

A separate switch allowed the S to abort a trial if he realized that his search had gone astray. This momentary pushbutton (U.L. #83050-1) was mounted vertically on a separate 3 x 3 inch (7.6 cm) response panel located on the table supporting the terminal. The "error" switch was placed next to the terminal on the side of the S's response hand. A size "large" Everlast 16-ounce boxing glove was worn on the S's idle hand to preclude use of two hands on the task.

**Procedure**

**Prestest.** All 42 Ss participated in the preliminary half-hour session in order to become familiar with the menu selection task and to provide baseline data for group assignment. Subjects were brought to the cubicle and were seated facing the experimental terminal. After making a brief explanation concerning the presentation of instructions, the experimenter (E) left the cubicle and closed the door. The S then read the first page of instructions displayed on the terminal, pressed the upper left pushbutton to advance to the second page, and continued to read at his own pace. The entire pretest instruction set can be found in Appendix B. The instructions referred S to a paper diagram on the desk depicting the
pretest word hierarchy. It consisted of 36 goal words broken into six categories of six words as shown in Figure 4. None of the goal words nor the six category descriptors were terms used later in the experiment. The configuration of $6^2$ was chosen such that C of 6 was not encountered by the Ss during later performance on the experimental or control tasks. Care was taken to insure that the words were very familiar and the categories obvious.

After referring to the word hierarchy, the instructions described the task and gave an example of a typical trial. At the end of the instructions, Ss had an option to begin the task or call the E for questioning by pressing the error button on the desk. Once started, the search trials were identical in timing to those of the experimental goal acquisition conditions. The terminal was configured such that display fields and response button positions occupied the lower three locations on each side of the terminal. Subjects performed 36 search trials, had a short break where E clarified any unclear aspects of the task, and finished the session with 36 more trials. The entire pretest session took approximately 30 minutes.

If the S was one of the first 21 Ss run, he was given a word diagram (chosen at random from $2^6$, $4^3$ and $8^2$ hierarchies) to take home and review prior to the return sessions. Subjects from the second half of the pretest sample were told they would receive a word diagram in the mail. Subjects were encouraged to review the word diagram so that they would understand the semantic organization of the assigned menu hierarchy. During preliminary testing, it was found that pilot Ss experienced some difficulty choosing appropriate categories from the upper-level menus unless they became familiar with the organization prior to the task.
Figure 4. Pretest word hierarchy
Ss were assigned to conditions by ranking all 42 on the basis of the median AT for the last 36 pretest trials, partitioning into seven strata, and selecting one S from each stratum to make up each group. Where this was not possible due to prior assignment, Ss were chosen to produce equivalent group means of pretest AT scores. After the pretest sessions and group assignments were completed, word diagrams were mailed to the second 21 Ss.

**Goal Acquisition Task.** Six Ss returned to participate in each of the four experimental conditions (E2^6, E4^3, E8^2, 64^1). Each S's two experimental sessions took place on consecutive days, anywhere from one to 25 days after the pretest session. After having fit the boxing glove to S's idle hand, E left the cubicle and initiated presentation of the self-paced instruction set on the S's display. Subjects were instructed to review the word diagram for (2.0 x total number of words in the hierarchy) seconds prior to beginning the goal acquisition trials. One sample trial was reviewed and Ss were reminded to perform the goal acquisition task as quickly as possible without making errors. Complete instructions can be found in Appendix B.

Each goal acquisition trial began with the presentation of a goal word in the center of the display. The order of presentation was determined by randomly selecting each goal word without replacement from the set of 64 goal words. After 2.0 seconds presentation the screen was blanked for 1.5 seconds, followed by presentation of the first level menu which remained until the S chose a category or timed out after 10 seconds. Following the S's response, the screen blanked for 0.5 seconds and then displayed the next menu called for by the S. This continued for each level of the hierarchy until the goal word was selected from
the lowest level menu. On correct trials the word "CORRECT" appeared along with the S's AT time rounded to the nearest 100 milliseconds. On incorrect trials the word "ERROR" was displayed. Either feedback message appeared 0.5 seconds after the final response and remained displayed for 2.0 seconds. A 4.0 second rest separated trials. The timing sequence for goal acquisition trials is summarized in Figure 5. Acquisition time was calculated as the sum of the individual response times necessary to acquire the goal word. Each partial response time (PRT) measured the time elapsed from menu presentation to S's response. When S hit the error button, having realized an erroneous response or having forgot the goal word, the "ERROR" message interrupted the trial and a new trial began after the 4.0 second rest. Error trials were repeated once at the end of each half-block of goal acquisition trials.

Each S performed two blocks of goal acquisition trials per session, taking about one hour. Each block consisted of one pass through the entire goal word set, or 64 trials. Subjects performed half-blocks (32 trials) at a time, separated by groups of 16 motor control (MC) trials. The MC task was included to get an estimate of the motor component of a PRT. Here, the computer selected one of the C display fields at random and presented a horizontal string of 12 contiguous X's at the chosen location. The Ss task was simply to press the corresponding pushbutton as quickly as possible. The timing sequence for the MC task is shown in Figure 6. Although they were encouraged during breaks to work "as quickly as possible without making errors," Ss were not informed of their performance relative to other participants at any time during the study. They were informed of the purpose of the study after their final session.
Figure 5. Flowchart of goal acquisition trial timing sequence (example shown for E4^3 condition)
Figure 6. Flowchart of motor control trial timing sequence (example shown for E43 and C43 conditions)
For each trial the computer recorded the goal word, its proper menu path through the hierarchy, the word path chosen by S, PRT for each level and AT in milliseconds. If S did not respond or caused a timeout to occur, the trial was coded with a Z and zeros occurred in place of PRT times for levels subsequent to the level of timeout. If S completed the trial but at some point made an incorrect selection the trial was coded with an "E" for error. If S hit the error button on the table, the trial record was coded with an "A" signifying abort.

**Control Tasks** - Subjects assigned to control conditions returned anywhere from one to 31 days after their pretest sessions. These Ss performed sub-tasks of the goal acquisition sequence identified as category matches (CM) and identity matches (IM) as well as the short blocks of MC trials.

**Category Match** - These trials presented the randomly-selected goal word followed by a higher level category menu chosen from the goal word's valid path. The level of abstraction was also chosen at random without replacement. The task was to select the appropriate category from the given menu as quickly as possible. Once a response was made, feedback identical to the goal acquisition trials was given. Thus each trial consisted of only one response indicating category choice. Timing was identical to the goal acquisition trials except that 2.0 seconds separated CM trials to conserve time. The number of CM trials necessary to equate practice with the experimental group varied with the depth of the hierarchy involved. For the $C_8^2$ condition, having only one category level, Ss performed 64 CM trials for each block of 64 goal acquisition trials. However, for the $C_2^6$ condition, having five category levels, Ss performed 320 CM trials for each block of practice.
For this reason, C2^6 Ss returned for four one-hour sessions, as only one block could be completed in an hour.

**Identity Match** - Completing the sub-tasks of the goal acquisition process is choosing the goal word from a menu containing it. Identity match trials presented the randomly-selected goal word followed by a menu from the same level containing the goal word. Since all of the hierarchies have only one goal word level, 64 IM trials constituted one block for all of the control conditions. Figure 2 summarizes the number and sequence of trials performed for each block of practice for each of the groups. Complete control group instructions can be found in Appendix B.
RESULTS

Pretest

Seven groups of six subjects were constructed using stratified samples based upon median pretest ATs from the second of two 36-trial blocks, which ranged from 1.65 to 3.21 seconds, with a mean of 2.15 seconds. The seven groups' mean of median AT scores ranged from 2.09 to 2.22 seconds. The groups' median ATs ranged from 1.98 to 2.16 seconds. Error rates corresponding to the AT scores ranged from 0.0 to 7.7 percent with a mean of 2.3 percent. A summary of pretest AT statistics can be found in Table 1.

The Pearson product-moment correlation relating pretest AT and the median AT for practice Block D within each experimental condition, was significant for the E2^6 and E4^3 conditions, \( r = .92, p < .01, r = .91, p < .05 \), but not for the E8^2 and 64^1 conditions. It should be noted that the precision of the \( r \) estimates is extremely low due to the small sample size of six Ss.

Goal Acquisition

Goal acquisition performance differed markedly among the four experimental conditions in terms of AT, commission of errors, variability, and learning. Figures 7, 8, 9, and 10 show each S's 32-trial median AT plotted over practice for the E2^6, E4^3, E8^2, and 64^1 conditions respectively. (Since Ss performed 32 trials without interruption, many analyses were done using 8 levels of practice each involving half of the
Table 1

Pretest AT Statistics for Seven Groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>Median</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>E26</td>
<td>2.22</td>
<td>2.06</td>
<td>1.33</td>
</tr>
<tr>
<td>E43</td>
<td>2.13</td>
<td>2.03</td>
<td>.96</td>
</tr>
<tr>
<td>E82</td>
<td>2.20</td>
<td>2.03</td>
<td>.82</td>
</tr>
<tr>
<td>C26</td>
<td>2.12</td>
<td>2.04</td>
<td>.63</td>
</tr>
<tr>
<td>C43</td>
<td>2.11</td>
<td>1.96</td>
<td>1.06</td>
</tr>
<tr>
<td>C82</td>
<td>2.09</td>
<td>1.98</td>
<td>.58</td>
</tr>
</tbody>
</table>
Figure 7. Median acquisition times for six subjects over practice in the E2$^6$ condition.
Figure 8. Median acquisition times for six subjects over practice in the E4³ condition.
Figure 9. Median acquisition times for six subjects over practice in the $E8^2$ condition.
Figure 10. Median acquisition times for six subjects over practice in the $64^1$ condition.
64 goal word set rather than the 4 based on completion of the 64-word set. To avoid confusion, analyses of the former type will refer to Blocks 1-8, and analyses of the latter type will refer to Blocks A-D.) Visual inspection of the figures yields some obvious differences among conditions. Groups E2^6 and 64^1 show more intersubject variability than E4^3 and 64^1. Most Ss, regardless of condition, reached asymptotic levels of AT in Blocks 5-8.

When group means of Ss' 32-trial median ATs are plotted, differences among the groups' AT performance become more obvious (see Figure 11). Groups E2^6 and 64^1 show higher AT scores than groups E4^3 and E8^2, with no overlap of curves. The fastest conditions, E8^2 and E4^3, demonstrate a difference at early levels of practice, but then converge to virtually no difference later in practice. A similar trend occurs between E2^6 and 64^1 except that the latter levels off by Block 4 and the former continues to decrease until Block 7. Asymptotic performance is evident in all four conditions by Block 7. Groups E2^6 and E4^3 again show more of a practice effect than E8^2 and 64^1.

A two-way analysis of variance (Structure x Practice) based on AT yielded significant main effects for structure, $F(3,20) = 21.4, p < .00005$, practice $F(7,140) = 14.0, p < .00005$, and their interaction $F(21,140) = 2.59, p < .0005$. A Newman-Keuls multiple contrasts test showed significant differences between the following contrasts at $p < .01$: E2^6 and E4^3, $Q(3,20) = 8.25$; E2^6 and E8^2, $Q(4,20) = 8.76$; 64^1 and E4^3, $Q(2,20) = 7.19$; 64^1 and E8^2, $Q(3,20) = 7.70$. Thus, the main effect of structure in the analysis of variance is due to the large discrepancy in AT between the two slowest conditions (E2^6 and 64^1) and the two fastest (E4^3 and E8^2). The significant interaction of structure
Figure 11. Mean acquisition times for each experimental group
and practice verifies what can be observed in Figure 11, that practice has a different effect upon the various structures.

The finding of two fast and two slow conditions suggests a U-shaped function of AT versus display choices. Figure 12 shows the overall mean AT for each group plotted against display choices for Blocks A-D. The ordinal relationship of conditions based on AT shifts with increased levels of practice. Also, the relative amounts of reduction in AT over practice can be seen in the vertical expansion or compression of the vertices in the plotted lines.

Error data were tabulated using the combination of misses, aborts, and time-outs in all conditions. Percent error was calculated by dividing combined errors by the total trials for that block and multiplying by 100. Figure 13 shows percent error over Blocks 1-8 for the four experimental conditions. The most error-free condition was E8² followed by 64¹. The highest error rates were found in E2⁶ and E4³ with the former higher than the latter in all but two blocks. Percent error decreased by a factor of two over the first four blocks for E2⁶ and E4³, whereas E8² remained stable and 64¹ increased, decreased and then kicked up again on the final 32 trials. It should be pointed out that percent error for 64¹ is probably inflated due to the 25-second time limit on the search task.

A two-factor analysis of variance (Structure x Practice) based on error rates demonstrated significant main effects for structure $F(3,20) = 6.86, p < .005,$ and practice, $F(7,140) = 3.37, p < .005,$ but not for their interaction. The Newman-Keuls multiple comparison method revealed significant differences in error rates for the following contrasts: $E2⁶$ and $E8²,$ $Q(4,20) = 5.71, p < .01; E4³$ and $E8²,$
Figure 12. Mean acquisition times for groups plotted against display choices for four levels of practice.
Figure 13. Mean percent errors for each experimental group
Since both ATs and percent error differed significantly over hierarchical structure, an examination of the relationship between the two dependent measures is appropriate. Figure 14 shows the overall means collapsed over practice for AT and percent error plotted adjacently in a bar chart. Ignoring the 64^1 condition for the moment, it appears that AT and error rate both decrease as display size (breadth) increases and number of levels (depth) decreases. The E8^2 condition allows fastest acquisition and fewest errors. There does seem to be a limit to this trend in that the 64^1 condition's AT and error rate both increase abruptly at the right-hand side of the chart.

Control Tasks

Data for the three control conditions C2^6, C4^3, and C8^2 consist of CM times, IM times and their respective error rates. Figures 15 and 16 show CM and IM times for the three conditions plotted over practice. Both CM and IM data curves show an increase in response time with an increase in the number of displayed items. The three CM curves descend with practice and are virtually parallel except for one overlap on Block 1. Category match times are consistently greater than their IM counterparts except for C8^2, where they are virtually equivalent. The IM curves show more spread on the time axis demonstrating a greater dependence upon the number of displayed choices. In addition, the IM curves appear more flat than the CM curves suggesting less of a practice effect.

The analysis of variance (Structure x Practice) based on CM data produced significant main effects for structure, $F(2,15) = 4.57$, $p < .03$, and practice, $F(7,105) = 19.25$, $p < .00005$, but failed to show...
Figure 14. Overall mean AT and percent errors for experimental groups
Figure 15. Mean category match times for each control group
Figure 16. Mean identity match times for each control group
a significant interaction at the alpha = .05 level. These results concur with the graphical analysis. A second analysis of variance for IM data produced similar results. The main effects of structure, $F(2,15) = 91.63, p < .00005$, and practice $F(7,105) = 3.63, p < .002$, were significant, but their interaction was not.

Several analyses were done comparing the control tasks with the analogous experimental PRTs. The CM scores for the $2^6$ hierarchy were sorted by level (five levels) and compared to the five levels of PRT scores from the experimental goal acquisition trials. A three-way analysis of variance (Experimental/Control $\times$ Levels $\times$ Practice) yielded no difference between conditions, but did produce significant differences among the five levels, $F(4,40) = 13.88, p < .00005$, and blocks of practice $F(3,30) = 11.75, p < .00005$. A similar analysis of variance for the $4^3$ hierarchy showed no levels or condition effects, but did demonstrate a significant practice effect, $F(3,30) = 28.39, p < .00005$. Comparisons between IM scores and the lowest level PRTs for the three hierarchies showed no differences between IMs and the lowest level PRTs for both the $4^3$ and $8^2$ hierarchies, but PRTs significantly slower than IMs, $F(1,9) = 15.74, p < .005$, for the $2^6$ hierarchy. No difference was found between IM and CM scores for the $C8^2$ condition.

Error rates were calculated for CM and IM trials using the same formula used for the AT errors. Percent error is plotted versus practice in Figure 17. Category match errors were greatest for the $C2^6$ condition, followed by $C4^3$ and $C8^2$. Identity match errors were much fewer than CM errors with no apparent ordinal relationship. Unlike the other conditions, the two types of error rates for $C8^2$ were very similar, both being less than 1.0 percent.
Figure 17. Mean percent errors for CM and IM tasks in control conditions.
A two-factor analysis of variance (Structure x Practice) based on CM error rates produced a significant main effect for structure, $F(2,15) = 6.95$, $p < .005$. Neither the main effect for practice, nor the structure x practice interaction, was significant at the .05 level. A similar analysis for IM error rates produced no differences. These results concur with what can be observed in Figure 17.

Two bar charts show the relationship between CM and IM times with their respective error rates in Figure 18. As was the case with the AT data, those plotted in Figure 18 represent group means over all levels of practice. The CM bar chart demonstrates that as CM times increase from C2$^6$ to C8$^2$, error rates decrease. The same trend does not hold for the IM data, however, the IM times increase a greater amount from C2$^6$ to C8$^2$, but the error rates do not decrease systematically.

Motor Control Task

Subjects in all seven conditions performed 16 motor control (MC) trials between blocks of AT or control task trials. Mean response times are plotted against practice for each group in Figure 19. As can be observed from the graph, the results are somewhat anomalous. Considering the fact that the MC task is identical for experimental and control conditions and differs only in the number of potential stimuli locations (C), the ordinal relationship of the data is curious. Looking at the experimental conditions only there exists an orderly decrease in MC time as C decreases for 64$^1$, 8$^2$ and 4$^3$ as well as a slight practice effect. However, MC data for C2$^6$ fall above the 8$^2$ curve and show no decrease of MC time with practice. The control group data for C4$^3$ and C8$^2$ lie near their experimental counterparts; however, they show less of a difference due to C. The C2$^6$ curve lies at the bottom of the figure.
Figure 18. Overall mean CM, IM and percent errors for control groups.
Figure 19. Mean motor control times for all seven conditions
where it would be expected. It should be noted that the E26 and E82 MC curves are based on four Ss' data rather than the usual six due to technical difficulties early in the study and one S who misinterpreted the MC task instructions. Error rates were not included in the analysis because they were negligible in each condition.

A three-way analysis of variance was performed on the six paired experimental and control conditions (excluding 641) to test for effects of structure, practice and the experimental/control factor. The main effect of structure was found not to be reliable at the .05 level. The main effect of practice was highly significant, \( F(7,182) = 12.03, p < .00005 \), and the main effect for experimental/control was also significant, \( F(1.26) = 5.33, p < .03 \). The only significant interaction was that for structure x experimental/control, \( F(2,26) = 5.48, p < .02 \). These results seem to be due to the anomalous MC times of the E26 group.

When a two-factor analysis of variance (Structure x Practice) is performed on the selected conditions of C26, E43, E82 and 641, the structure's main effect becomes highly significant, \( F(3,18) = 16.82, p < .00005 \). The practice effect remains highly significant, \( F(7,126) = 7.92, p < .00005 \), but no interaction was found. These results agree with the graphical analysis by proving parallelism over practice and a structure effect for the selected groups.

**Model Prediction of Goal Acquisition Performance**

An additive model was proposed to account for AT by summing the constituent CM and IM components. A prediction was made for ATs using the proposed additive model and data derived from previous visual search literature. The resultant predictive curve is plotted with the overall mean AT results in Figure 20. Although the predictions overestimate AT
Figure 20. Overall mean acquisition times (OMAT) and original predicted acquisition times (OPAT) vs. display choices.
in all but one case and the ordinal relationship of $2^6$ and $64^1$ are reversed, there appears to be a reasonable fit in terms of shape and absolute values. A multiple t-test yielded a significant difference for the $2^6$ comparison alone, $t(5) = 3.10, p < .05$, demonstrating that the estimate for AT is lower than the overall observed mean AT for the $2^6$ structure. However, for the last block of practice, the group mean AT for $E2^6$ dropped to 4.81 seconds which fails to be significantly differently from the predicted value of 4.32 seconds at the .05 alpha level.

A more reasonable test of the additive model is to derive CM(C) and IM(C) functions from the control data, and to use them to estimate the constituent additive components of acquisition time. In order to derive estimates of CM(C) and IM(C), individual Sa's CM and IM data were used in simple linear regression analyses to find the best fitting straight lines for the four different levels of practice. The resultant regression parameters can be found in Table 2. A reasonably high degree of fit ($R^2 > .85$) and consistent, significant slopes were found for the IM data. The CM parameters indicate poor goodness-of-fit ($R^2 < .58$) and a steepening slope over practice. Intercepts for both functions decrease over practice, but the decrease is more pronounced for the CM data.

When the regression line estimates for CM(C) and IM(C) are used as input to the additive model, predicted AT scores ($\hat{AT}$) can be calculated for each structure using equation (1):

$$\hat{AT}(L,C) = \sum_{n=1}^{L-1} CM_n(C) + IM(C)$$

Predicted values for the $64^1$ condition were calculated by extrapolating the IM(C) curve to IM(64). These predictions are compared to the
Table 2
Regression Parameters for CM(C) and IM(C)

<table>
<thead>
<tr>
<th>Task</th>
<th>Block</th>
<th>Intercept</th>
<th>Slope</th>
<th>p</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td>1.076</td>
<td>.016</td>
<td>.35</td>
<td>.06</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>.909</td>
<td>.034</td>
<td>.04</td>
<td>.24</td>
</tr>
<tr>
<td>CM</td>
<td>C</td>
<td>.781</td>
<td>.035</td>
<td>.003</td>
<td>.44</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>.712</td>
<td>.050</td>
<td>.0002</td>
<td>.58</td>
</tr>
<tr>
<td>A</td>
<td></td>
<td>.396</td>
<td>.107</td>
<td>.0001</td>
<td>.86</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>.310</td>
<td>.117</td>
<td>.0001</td>
<td>.88</td>
</tr>
<tr>
<td>IM</td>
<td>C</td>
<td>.288</td>
<td>.111</td>
<td>.0001</td>
<td>.87</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>.300</td>
<td>.108</td>
<td>.0001</td>
<td>.87</td>
</tr>
</tbody>
</table>
actual ATs for four levels of practice in Table 3. As the table indicates, only three ATs significantly differ from the observed AT values, $t(5) > 2.57, p < .05$. All of these are cases where IM(64) is used to estimate search time for the 64-word menu. The remaining AT for 64 shows a large discrepancy as well, but fails significance by a small amount $t(5) = 2.52$. The comparisons for the E8 condition produced rather large $t$ values as well, probably due to the small standard error terms for that condition. The additive model made consistent, accurate predictions for E2 and E4 as evidenced by small $t$ values over all four blocks of practice.

Each t-test was performed by comparing the group mean based on six Ss with one calculated AT value. In order to give each t-test a fair chance of being significant, alpha was set at the .05 level. This raises the experiment-wise alpha level to .56. This means that the probability that at least one of the 16 tests is spuriously significant is between .05 and .56, presuming that all 16 null hypotheses are true.

Figures 21 through 24 show the regression lines for CM and IM, the resultant AT curve, and the observed AT curve for practice Blocks A through D. Disregarding the 64 points, the apparent fit improves with increasing levels of practice. In fact, for all blocks, AT falls within the 95 percent confidence interval for these AT means.
### Table 3
Comparisons of Mean AT with Predicted AT

<table>
<thead>
<tr>
<th>Group</th>
<th>Actual AT</th>
<th>Predicted AT</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>E26</td>
<td>7.039</td>
<td>6.156</td>
<td>1.42</td>
</tr>
<tr>
<td>E43</td>
<td>3.389</td>
<td>3.108</td>
<td>.76</td>
</tr>
<tr>
<td>BLOCK A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E82</td>
<td>2.755</td>
<td>2.459</td>
<td>2.23</td>
</tr>
<tr>
<td>641</td>
<td>5.707</td>
<td>7.233</td>
<td>2.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E26</td>
<td>5.687</td>
<td>5.434</td>
<td>.57</td>
</tr>
<tr>
<td>E43</td>
<td>2.728</td>
<td>2.871</td>
<td>.59</td>
</tr>
<tr>
<td>BLOCK B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E82</td>
<td>2.643</td>
<td>2.429</td>
<td>1.79</td>
</tr>
<tr>
<td>641</td>
<td>5.053</td>
<td>7.779</td>
<td>10.20*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E26</td>
<td>5.205</td>
<td>4.767</td>
<td>.89</td>
</tr>
<tr>
<td>E43</td>
<td>2.552</td>
<td>2.576</td>
<td>.10</td>
</tr>
<tr>
<td>BLOCK C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E82</td>
<td>2.541</td>
<td>2.241</td>
<td>2.51</td>
</tr>
<tr>
<td>641</td>
<td>5.320</td>
<td>7.417</td>
<td>3.89*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E26</td>
<td>4.810</td>
<td>4.574</td>
<td>.54</td>
</tr>
<tr>
<td>E43</td>
<td>2.434</td>
<td>2.554</td>
<td>.63</td>
</tr>
<tr>
<td>BLOCK D</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E82</td>
<td>2.471</td>
<td>2.272</td>
<td>2.01</td>
</tr>
<tr>
<td>641</td>
<td>4.942</td>
<td>7.186</td>
<td>3.90*</td>
</tr>
</tbody>
</table>

* significant at p < .05  Critical value t(5).05 = 2.57
Figure 21. Observed AT and predicted AT with CM and IM regression lines: Block A
Figure 22. Observed AT and predicted AT with CM and IM regression lines: Block B
Figure 23. Observed AT and predicted AT with CM and IM regression lines: Block C
Figure 24. Observed AT and predicted AT with CM and IM regression lines: Block D
DISCUSSION

Pretest

Every S in the experiment found the pretest task easy to perform, and several commented that it served as a good introduction to the goal acquisition procedure. The range of pretest ATs was surprisingly large. The slowest S's median AT of 3.21 seconds is almost twice that of the fastest S (1.65 seconds). Given this level of individual differences, the decision to preselect group composition on pretest performance appears to have been wise. Balancing groups was hampered somewhat due to attrition which caused the control groups mean protest ATs to be slightly lower than their experimental counterparts. This difference may have caused estimates for $\hat{AT}$ using control Ss data as input to the additive model to be predominantly lower than the observed AT values. Table 3 shows that for conditions $E2^6$, $E4^3$, and $E8^2$, $\hat{AT}$ falls below AT 9 out of 12 times.

Pretest performance correlated positively with subsequent goal acquisition performance for $E2^6$ and $E4^3$ Ss, adding validity to the pretest selection procedure. Lack of correlation in the $64^1$ group is not surprising, considering that the task involved a single-display search for the goal word. However, one would expect correlation between pretest and $E8^2$ performance since it is the goal acquisition task most similar to the pretest. Although the $r$ values never attained significance for this group, they ranged from +.57 to +.78 over four blocks of
practice. These $r$ values are in the correct direction, and quite respectable but fail to be conclusive with only six pairs of observations in each test. By comparison, the $r$ values for the 64\textsuperscript{1} condition were $+.17$, $-.17$, $+.53$ and $+.66$, indicating much less of a relationship.

In retrospect, the pretest task appears to have been worth the time and effort. In addition to teaching the goal acquisition task and providing group selection criteria, it provided another combination of $C$ and $L$ by which the additive model could be tested. The pretest also provided a mechanism by which unreliable or visually handicapped $S$s could be screened from the study. No candidates were actually dropped for visual problems (though some were, due to tardiness), however two $S$s realized the need for their corrective lenses for the return test sessions.

**Goal Acquisition**

All experimental $S$s found the goal acquisition task to be entertaining and challenging. Most of them also appeared to be motivated to win the bonus and appreciated the trial-by-trial feedback. Wearing the boxing glove on the idle hand offended several of the $S$s' integrity, but was explained as an experimental control necessary to support the credibility of the data.

The differences found among the four experimental conditions in terms of variability, learning, errors and AT support the original hypothesis that hierarchical organization affects goal acquisition performance in a menu-based information system. The results also demonstrate the validity of the notion of an optimal tradeoff between levels of indentation (system depth), and menu choices (system breadth). The hypothesized predictive additive model was found to be accurate in predicting and
describing observed AT data for the present system. It will be discussed in a subsequent section.

One difference observed among the experimental groups is inter-subject variability. It appears from Figures 7 through 10 that group variability was greatest in conditions with the slowest ATs (E26 and 641). Groups having shorter ATs (E43 and E82) had less spread among the six group members. Upon inspection, group variability in pretest scores does not correlate with AT variance. It would be fair to conclude that the differences in AT variance result from differences inherent in the hierarchical organizations. Intersubject variability may not necessarily be a drawback for a given system configuration, unless training and personnel selection become important considerations. Obviously, E82 is superior in regard to intersubject variability, followed by E43.

Within-S variability is approximately equal across conditions, except for 641 where S's vary as much as 4.0 seconds between blocks. The source of this variance is quite obvious. If the goal word happens to be located near the initial scanning location the word will be found quickly. When located near the end of the scanning pattern, retrieval will be very slow. This kind of uncertainty in AT can have profound adverse effect upon system performance in applications stressing contingency or dependence among system components. For instance, if a complex Command, Control and Communication system depends on several sources of information to make a tactical decision within a time constraint, the system must be confident in its sources' abilities to provide input within specific time parameters. If source delays are too long, the system can execute a more simple decision and retain some minimal probability of success. If the system cannot predict whether it
will be served by its sources in time to execute the complex decision, it will be unable to make any decision. Thus, in some cases it is far better to be reliably slow than unreliably fast. A long-listed menu system would definitely be a hindrance in such a system.

The four experimental conditions also demonstrated differences in the effect of practice on AT. The E2\textsuperscript{6} condition showed the greatest reduction in AT over practice, and E\textsuperscript{2} the least. The E4\textsuperscript{3} condition fell somewhere in between, which suggests that more learning is possible with a greater number of levels and categories. As levels increase, the number of item-category relationships increases, and these have to be learned in order to do the goal acquisition task efficiently. The two conditions with the fewest levels and item-category relationships (E8\textsuperscript{2} and 64\textsuperscript{1}) show the least learning. The slight learning effect for 64\textsuperscript{1} probably is attributable to Ss gaining familiarity with the words and developing search strategies which key on physical aspects of the words.

Interacting with the entire goal word set four times has evidently allowed learning to reach asymptotic levels. The practice necessary for asymptotic learning increases with the number of menu levels in the hierarchy. The E8\textsuperscript{2} Ss' performance leveled off almost immediately, making it the best for applications where unskilled operators would not receive much training. The E2\textsuperscript{6} Ss' learning continued until the seventh block. Considering the fact that practice was extremely concentrated in the experiment, this extended learning time could translate into days or weeks in a less concentrated form. Therefore, in applications where training is minimal or system use is intermittent, deep menu structures could be a problem.
The speed at which goal items could be acquired varied dramatically across the four experimental conditions. Conditions E8^2 and E4^3 were the fastest, with almost identical group means after Block 3. Conditions E2^6 and 64^1 were approximately twice as slow, with no reliable difference between them. Looking at Figure 11, 64^1 was faster than E2^6 in the early stages of practice, Q (2,20) = 2.99, p < .05. Likewise, E8^2 was quicker than E4^3 before asymptotic performance was reached, but the difference was not significant at the .05 level. Therefore, in applications where speed is required with little or no training, the order of preference for structure is 8^2, 4^3, 64^1 and 2^6. When plotted against display choices, the resultant U-shaped function of AT verifies the notion of an optimum tradeoff between depth and breadth for a system of this size.

Goal acquisition speed is not the only factor to be considered when evaluating performance. Errors can have costly ramifications upon system effectiveness. Structure did have a significant impact on error rates. The conditions with the fewest levels had the fewest errors. The E8^2 condition had fewer errors than 64^1. This was probably a result of the 25-second timeout for Ss in the 64^1 task. Time-out trials were combined with error trials thereby inflating the error rate. The time-out feature was implemented to motivate Ss' search through the 64-word menu. It was never anticipated that Ss would fail to locate goal words in this amount of time.

It was hoped that the instructions to "find the goal words as quickly as possible without making any errors" would equalize error rates across conditions, thereby avoiding speed-accuracy tradeoff. As it turned out, error rates were significantly different, but increased
with AT. The bar chart in Figure 14 shows that for all but the 64\(^1\) condition, errors committed decreases with AT as number of levels decreases. The figure also punctuates the fact that the 8\(^2\) condition is best in terms of AT and errors. The chart also suggests that the minimum for error rate lags that for AT on the C continuum. Figure 25 presents combined smooth curve plots of AT, percent error and their sum against C and L. Although the C scale is compressed on the right it is clear that error rate reaches a minimum at a higher C value than AT. Errors reach a minimum near C = 8 while ATs bottom out between C = 5.5 and C = 6.0. If it is believed that the 64\(^1\) error rate is artificially inflated due to the time limit on the search task, then errors probably truly approach an asymptotic low level at some point C > 8. Their combined effect with equal weighting (AT + E) reveals a curve almost V-shaped with its vertex at or near C = 8. On the other hand, when these variables are graphed as a function of L, the number of hierarchical levels, the resulting smooth curves are quite different. The AT curve reaches minimum at L = 2.4 and errors are fewest at L = 1.95. The sum of AT and percentage error takes a fast dip at L = 2, sharply climbs up at L = 3 and levels off out near L = 6. The fact that these curves apply only to a system of 64 items must not be overlooked. Both sets of curves have a hidden third parameter which must be considered upon interpretation. In fact, the AT minima of C = 5.7 and L = 2.4 would produce a system the size of \(C^L = 5.72.4 = 65.2\) items, demonstrating the hidden constraint \(C^L = 64\) for all points on the AT curves in these graphical analyses.

In summary, it would be fair to conclude that for a system of moderate size the number of hierarchical levels should be minimized, but
Figure 25. Acquisition time, percent errors and their sum vs. display choices and levels
not at the expense of display crowding, for optimal goal acquisition performance. This guideline assumes, of course, that the semantic relationships allow such a broad organization. The data support the conclusion such that a menu hierarchy of two levels was the fastest, produced the fewest errors, showed the least variability, and was the easiest to learn. If for some reason two levels and eight choices per level cannot satisfy system requirements, expansion in breadth is recommended over expansion in depth.

Chances are, most information systems would require asymmetric organization due to the differential sizes of component categories. For instance, a department store inventory would have many more categories and items dealing with clothing than with photographic equipment. The menu organization designed to handle this situation would need a deeper structure under "clothing" than "photography." One method of limiting the levels necessary for the clothing inventory would be to sub-categorize "clothing" into "men's clothing," "women's clothing," etc. at the first level menu.

It should be emphasized that one of the assumptions made at the outset of the present investigation was that the task being studied is a speeded performance task. The recommendations or guidelines that result from the present data may not be ideal for applications where the speed of goal acquisition is not critical. Therefore the recommendations developed here may be more applicable to a military cockpit situation where speed determines survivability than to an inventory control problem in a small business. However, regardless of speed criticality the present results should provide useful guidelines for any application involving a hierarchical system of comparable size.
Control Tasks

Most of the Ss in the three control conditions found the CM and IM tasks boring and tedious. A few motivation problems developed when Ss realized the limited effect practice and effort had on trial times. As can be seen in Figure 15 the greatest practice effect was on the order of .40 seconds (C4^3 - CM) for any given group over eight blocks of trials. The trial feedback was given in tenths of a second, causing smaller differences in performance to remain obscure to the S. This level of feedback accuracy may be ideal for the goal acquisition tasks which exhibit a larger variance of scores but may cause frustration in the control tasks. Considerably less learning was exhibited by the group IM data. This seems logical in that the CM task requires the learning of an association between an item and its category whereas, improvement in the IM task is most likely a result of becoming familiar with the physical features of the goal words. All three curves in Figure 15 show evidence of asymptotic performance, lending credence to the intention of using control data to predict goal acquisition performance.

The relative positions of the three CM and the three IM curves demonstrate the effect of display choices (C), which proved to be statistically reliable. As was expected, IM and CM time both increase as a function of C. Regression analyses on the 18 S's data bear this out, with the slope for IM data greater than that for CM data. The lack of significant interaction between structure and practice for both IM and CM implies that there is no differential learning effect across the three conditions.

The error data for the control groups, unlike their experimental counterparts, showed errors to be constant over practice for both the CM
and IM tasks. This is surprising considering that there was a strong practice effect for the E2^6 and E4^3 conditions. However, inspection of Figure 17 reveals that the overall percent errors for CM in C2^6 is very near the final level of error commission in E2^6. Likewise, percent error for CM in C8^2 is nearly identical to that of E8^2. A similar relationship for CM errors and AT errors does not exist for the 4^3 hierarchy, however. The overall low, consistent errors for the IM tasks supports the interpretation made earlier that errors are more than likely attributable to item-category matching faults than item-item matching faults.

The bar charts in Figure 18 clarify the relationship between CM and IM times with errors. For the CM task, errors decrease systematically as CM times increase over C. This decrease is more likely a result of the concomitant decrease in levels and item-category relationships. No similar trend exists for the IM data. The IM error rates were all near the 1.0 percent level. Identity match errors probably resulted from Ss forgetting the goal word or hitting the wrong pushbutton accidentally.

**Motor Control Task**

Subjects commented that the MC task served as an effective break between blocks of word processing trials. There was a significant trials effect, suggesting that the motor skills involved in selecting menu choices by pushbutton improve with practice. This is consistent with findings in the field of motor control theory and skills learning.

There does seem to be a structure effect on MC times if the anomalous E2^6 condition is excluded from analysis. The effect can be attributed to differences in C only, as there were no other independent variables involved. The anomalous result of the E2^6 MC curve was traced to the fact that only four Ss' data were used in the analysis, thereby
exaggerating the slowest S's influence upon the group mean. This particular S was also 42nd out of 42 Ss on the pretest task.

The main objective in collecting MC data was to estimate the non-cognitive component of goal acquisition time. This can be done by multiplying the motor component time by the number of responses required and dividing by the total goal acquisition time for any response configuration. This results in estimates of .72, .58, .46, and .14 for the \( E_2^6, E_4^3, E_8^2 \) and \( 64^1 \) conditions respectively for the overall mean data across practice. The proportional increase of response component with number of responses may seem surprising. However, because less time is spent finding the proper menu item as \( C \) decreased (see IM and CM as functions of \( C \)), the response component dominates each response, driving up the proportion as \( C \) decreases (or \( L \) increases). This fact may be considered a further detriment to deep hierarchical systems in that the operator spends a greater proportion of his time making inputs to the machine and less finding the target goal item.

The MC data for the control tasks do not coincide with the intercepts of the CM(C) and IM(C) regression lines. If this had occurred, the MC time could be considered as a constant additive component of the IM or CM task and perhaps would have germinated an additive model based on three terms. Motor control times are .07 to .30 seconds higher than the IM intercepts and .125 to .60 seconds lower than the CM intercepts. This may indicate the MC task is a realistic estimate of the time necessary to encode the stimulus and respond, but cannot be used in a deductive way to estimate the purely cognitive component times. In order to propose a subtractive logic the assumptions of independent and non-overlapping processes must be made. Clearly, the motoric task of
responding to a pushbutton can physically overlap with the search task. The advantage of this overlapping during a display search would be to move the hand toward the remaining side of the display as the alternatives on the side searched first are eliminated. No analogous anticipation is possible during the MC task. This would tend to make the MC times appear greater than the component of the search task it was intended to model.

When overall mean MC times collapsed over practice are plotted against display choices (C) the result is not a linear function. However, when plotted against \( \log_2 C \), a linear function obtains, except for the \( E2^6 \) point (see Figure 26). The term \( \log_2 C \) represents the stimulus uncertainty in bits of information. For every doubling of the number of potential stimulus locations, the stimulus uncertainty increases by one bit. Therefore as \( C \) increases from 2 to 64, the stimulus uncertainty increases from 1 to 6 bits. This linear relationship between MC and stimulus uncertainty is no surprise. During the early 1950's several researchers observed the proportional increase in RT with stimulus information (Hick, 1952; Hyman, 1953). Thus, the MC data conform to the Hick-Hyman Law \( (RT = a + bH_S, \text{ where } H_S \text{ is the stimulus information}) \).

The general lack of satisfaction with the MC data may stem from two experimental oversights. First of all, the response configuration for the \( 64^1 \) condition was a compromise to physical limitations of the experimental apparatus. Sixty-four response buttons (one for each word displayed) would have been the logical extrapolation of the S-R pairing scheme used in the other conditions. The fact that only eight were used probably caused the CM times (as well as ATs) to be artificially short.
Figure 26. Motor control group means as a function of stimulus uncertainty
Secondly, in an effort to conserve session time, no feedback was given on the MC trials. There is no way to account for the possible effect this may have had on MC performance. Future efforts should benefit from the realization of these errors.

**Model Prediction of Goal Acquisition Performance**

The proposed model, which describes the goal acquisition times as a sum of CM times plus one IM time, predicted the observed ATs relatively well using data from the literature. If the current research were not possible, the estimates would have served as good guesses or "ball park" numbers for the various tasks. These numbers might indeed suffice for the purpose of a singular design application with wide performance tolerances. However, the objectives of this study include analysis, description and prediction. It is important to gain an understanding of how the task is performed, to describe its components and interactive variables in a comprehensible model, and to test the model's ability to forecast future outcomes.

Recognizing the fallibility of using someone else's data to predict experimental results, the control conditions were run to provide accurate input to the additive model. As was already demonstrated, the model predicts ATs very well using the CM(C) and IM(C) functions derived from the control data. The predicted AT values for the 64\(^1\) condition are the only failures of the model. These AT points are calculated by linear extrapolation of the IM(C) function to the point where \(C = 64\). This may be an inappropriate method in light of the previous comments on apparatus limitations. Also, one must consider the density with which the 64 words appeared on the screen. A true extrapolation of the C parameter would have required a display screen with eight times the
display area to maintain consistent spacing between words. Certainly, scanning a 25.4 x 19.8 inch (.50 x .64 m) display and finding the correct response button among 64 would have increased ATs for the 64\(^1\) condition. Considering these distortions, the \(\hat{AT}\) for 64\(^1\) seems to be quite good. Concentrating on the three other configurations for the moment I would like to discuss the CM(C) and IM(C) functions which provide the additive model with the information necessary to estimate AT.

The \(\hat{IM}(C)\) function was predicted at the outset with more confidence than the \(\hat{CM}(C)\) function. This was because the parent data were actual numbers taken from previous experiments whereas the original prediction of the \(\hat{CM}(C)\) function was an outgrowth of the \(\hat{IM}(C)\) function with minor changes in slope and intercept. The IM(C) function (generalizing across blocks of practice) attained from three control groups actually came very close to \(\hat{IM}(C)\).

\[
\hat{IM}(C) = .40 + .08 (C)
\]

\[
IM(C) = .32 + .11 (C)
\]

The IM(C) function also demonstrated good linear qualities. The four regression lines had consistent slope estimates (.107 to .117), all of which were significantly different from zero (\(p < .0001\)). The \(R^2\) values ranged from .86 to .87 demonstrating very high goodness-of-fit. The observed lower intercept of IM(C) may be due to good S-R compatibility of the task. The greater slope may be caused by the wider spacing of words on the display than that found in the printed word lists used in previous studies.

On the other hand, the CM(C) curve turned out to be far different from what was expected.
\[
\hat{C}(C) = .40 + .17 (C)
\]
\[
C(M) = .80 + .035 (C)
\]
The slope of CM(C) was not significantly different from zero until the second block of practice. In fact, the CM(C) curve appears to be logarithmic in early stages of practice. The slope becomes more steep with practice, and the low R^2 values demonstrate poor goodness-of-fit. Furthermore, the CM(8) data points do not appear any different from the IM(8) points. These developments are very discrepant from what was predicted.

When the mean data for CM and IM collapsed over practice are plotted against C with the response time scale expanded, the differences between the two functions become more obvious (see Figure 27). The slope of the IM line is obviously greater than the slope of the CM line. The CM line also has a slight bend downward above C = 4 while the IM line is virtually straight. Although there is no statistical difference, the CM point is lower than the IM point at C = 8.

One likely explanation for the CM(C) function is that since the control conditions were not orthogonal tests of two independent parameters, interaction of C and L produced attenuation of CM(C) at the C = 4 and C = 8 levels. Since C^L = 64 is a constraint on all three conditions, the C_2^6 Ss had 62 categories to learn, whereas the C_4^3 Ss had only 20, and those in C_8^2 had only 8. Since the level from which the CM trial was chosen was random, the uncertainty of the category stimulus was greater in the C_2^6 condition than for the C_4^3 and C_8^2 conditions. As Ss practiced, the CM task, they began to anticipate the choice of categories to be presented. This is probably why CM(8) is no greater than IM(8). With only eight categories to be learned, anticipation turns
Figure 27. Category match (CM), identity match (IM), and motor control (MC), group means as a function of display choices.
a CM into an IM. These observations are consistent with the notion that
the CM function is influenced not only by C, but also by the size of
the structure and its concomitant number of categories.

If an orthogonal set of control conditions were run such that L had
no effect on CM, the resultant CM(C) function probably would have been
more similar to IM(C) predicted. However, it would no longer predict
AT via the additive model. Further research employing orthogonal con­
trol conditions spanning the variables of C and L are necessary to further
analyze the nature of the CM task.

Alternative models were considered for comparison with the proposed
additive model involving CM(C) and IM(C) components (equation 1).
Predictions of AT based on MC times alone were calculated by multiplying
MC(C) by the required number of responses (L) for each hierarchy:

\[ \hat{AT}(L,C) = L \times MC(C) \]  

(4)

Similarly, predictions of AT were made using IM time estimates from the
fitted regression line:

\[ \hat{AT}(L,C) = L \times IM(C) \]  

(5)

An analogous simple model using only CM estimates took the following
form:

\[ \hat{AT}(L,C) = L \times CM(C) \]  

(6)

The resulting point estimates for the four hierarchies using equations
(4), (5) and (6) are listed in Table 4. Also listed are the overall
mean observed ATs, estimates of AT based on the additive model
Table 4
Comparison of Alternative Models Predicting AT

<table>
<thead>
<tr>
<th>Model/Data</th>
<th>( \hat{\theta}^6 )</th>
<th>( \hat{\theta}^3 )</th>
<th>( \hat{\theta}^2 )</th>
<th>( \hat{\theta}^1 )</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed AT</td>
<td>5.68</td>
<td>2.77</td>
<td>2.60</td>
<td>5.25</td>
<td>-</td>
</tr>
<tr>
<td>Additive Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- regression lines</td>
<td>4.89</td>
<td>2.64</td>
<td>2.28</td>
<td>7.36</td>
<td>0.78</td>
</tr>
<tr>
<td>- point data</td>
<td>5.14</td>
<td>2.85</td>
<td>2.45</td>
<td>7.36</td>
<td>0.67</td>
</tr>
<tr>
<td>- previous data</td>
<td>4.26</td>
<td>2.88</td>
<td>2.80</td>
<td>5.52</td>
<td>0.51</td>
</tr>
<tr>
<td>CM(C) x L</td>
<td>5.22</td>
<td>2.82</td>
<td>2.16</td>
<td>3.04</td>
<td>1.74</td>
</tr>
<tr>
<td>IM(C) x L</td>
<td>3.24</td>
<td>2.28</td>
<td>2.40</td>
<td>7.36</td>
<td>2.66</td>
</tr>
<tr>
<td>MC(C) x L</td>
<td>4.11</td>
<td>1.60</td>
<td>1.19</td>
<td>0.72</td>
<td>31.69</td>
</tr>
</tbody>
</table>
(equation 1) using point data rather than the fitted regression lines for CM(C) and IM(C), and the original predictions of AT using data from previous search studies. Chi-square statistics were calculated (df = 3) for the above models comparing the model predictions of AT with the observed data. Considered as relative estimates of goodness-of-fit (Hoel, Port and Stone, 1971), the Chi-square values can also be found in Table 4.

Surprisingly, the original estimates using previous visual search data provide the best fit to the observed AT data. This apparent accuracy is due mostly to the accurate estimate of AT for the 64 hierarchy. The next best fit is attained by the additive model (equation 1) using point data, followed closely by the same model using regression lines for IM(C) and CM(C). This slight advantage for the point data indicates some error inherent in the regression approximations of IM(C) and CM(C). The models based on IM and CM alone (equations 5 and 6) are relatively poor in estimating the AT data, followed by the model involving only MC(C) (equation 4) which is by far the poorest.

Figure 28 demonstrates three of the models graphically. The original additive model (equation 1) fits the observed data very well, except for the C = 64 point. The simple model using the IM(C) regression line estimates (equation 5) provides poor fit at low levels of C, improves at C = 8 where CM = IM, and over-estimates AT at C = 64. The model based on MC(C) alone provides a monotonically decreasing function as C increases, and the number of responses decreases. Thus the additive model using CM(C) and IM(C) components (equation 1) provides the best prediction of the observed ATs.
Figure 28. Various model predictions of acquisition time
Assuming that the control tasks are valid sub-tasks of the goal acquisition task, their respective error rates should predict error rates for the goal acquisition task. Error predictions based on control data were calculated by multiplying the probabilities of being correct on the various sub-tasks and subtracting the product from 1.0. This calculation assumes the sub-tasks are independent processes and that errors are equally probable at all levels of the goal acquisition task. Table 5 summarizes the observed error rates and the predictions based on the control data. Predictions for the 4^3 and 8^2 are quite good, however that for the 2^6 hierarchy is much greater than the observed error rate. This discrepancy may be due to breakdown of the above assumptions.

Conclusions

In conclusion, the present study has contributed to the understanding of how hierarchical organization affects goal item acquisition performance. The most salient findings were:

1) The structure of menu organization has a profound effect on acquisition time and the probability of error in a speeded performance context.

2) The concept of an optimal tradeoff between depth and breadth is valid for a system of given size.

3) The proposed additive model adequately describes the goal acquisition process and can be used to predict acquisition times using CM(C) and IM(C) functions.

4) For systems of moderate size system breadth is preferable to system depth.
Table 5

Goal Acquisition Error Predictions Based on Control Errors

<table>
<thead>
<tr>
<th>Hierarchy</th>
<th>CM</th>
<th>IM</th>
<th>Predicted</th>
<th>Goal Acquisition</th>
</tr>
</thead>
<tbody>
<tr>
<td>2^6</td>
<td>5.7</td>
<td>0.8</td>
<td>25.9</td>
<td>7.7</td>
</tr>
<tr>
<td>4^3</td>
<td>3.3</td>
<td>1.2</td>
<td>7.6</td>
<td>6.7</td>
</tr>
<tr>
<td>8^2</td>
<td>0.7</td>
<td>0.3</td>
<td>1.0</td>
<td>0.7</td>
</tr>
</tbody>
</table>
Future Research

Future investigative efforts in this area should consider several potential factors affecting goal item acquisition that were not addressed in the present study. First, the effect of overall system size upon the depth-breadth tradeoff should be studied. Secondly, the effect of consistent placement of menu choices within a given menu and the resultant use of location cues by the operator should be related to the present findings. Although most computer menu systems implement consistent menu choice locations, the underlying assumptions made for the present methodology precluded such an arrangement. Thirdly, the nature of the CM component of the goal acquisition task should be further investigated to determine what variables affect time to make a category match.

It is hoped that this study may serve as a catalyst for continued investigation of the way hierarchical system organization influences the acquisition of system elements. The notion of the depth-breadth tradeoff does not apply only to computer menu interaction, but pervades keyboard design, conceptualization of supervisory control of large systems and the interaction with system failure diagnostics as well.
APPENDIX A

Word Hierarchies
Figure 29. Word hierarchy for $E_8^2$ and $C_8^2$ conditions.
Figure 30. Word hierarchy for $E4^3$ and $C4^3$ conditions
Figure 31. Word hierarchy for E2^6 and C2^6 conditions
Figure 1 (continued)

COUNTRY
  EUROPE
    EASTERN
    WESTERN
  ASIA
    NORTHERN
    SOUTHERN

LAND
  DESERT
  MOUNTAIN
  EVEREST
  MOUNTAIN

WATER
  OCEAN
  ATLAN TIC
  PACIFIC
  LAKE
  HURON
  ERIE

CULTURE
  ART
    MUSIC
      OPERA
      SYMPHONY
    PERFORMANCE
      DANCE
      BALLET
      WALTZ
    GRAPHICS
      DRAWING
      PAINTING
    DECORATION
    SCULPTURE
      POTTERY
      CARVING

SOCIETY
  PERSON
    MODERN
      PRESIDENT
      LINCOLN
      ROOSEVELT
    GENERAL
      PATTON
      MACARTHUR
    RULER
      CLEOPATRA
      CAESAR
    ANCIENT
      PHILOSOPHER
      PLATO
      ARISTOTLE
APPENDIX B

Instructions to Subjects
INSTRUCTION SET FOR THE PRETEST

You are participating in a study investigating how menus should be organized for computer-based systems. Today's session will take no longer than 1/2 hour. Two future sessions will take approximately 1 hour each. You will be paid at the end of the third session for your participation. You will receive a fixed rate of $3.00 for each session as well as a bonus of $5.00 if you are best in your group. Thank you for helping with this research. I think you will find it fun and challenging.

(Press upper left button to display next page of instructions.)

******************************************************************************

In order for you to be able to do today's task, please turn over the sheet marked "P" and study the groups of words for about a minute and a half.

******************************************************************************

Now that you have examined the 6 categories of words, the task should be easy to follow:

The computer will pick one of the 36 words from the list at random and present it on the screen. Commit this "goal word" to memory. The goal word will then disappear.

******************************************************************************
Shortly after, you will see a list of 6 category names on the screen. These 6 categories will be those found in the boxes on the diagram sheet. Your task begins by picking the appropriate category for the goal word memorized. Simply press the button next to the correct category word.

******************************************************

After selecting the category word, a second list of 6 words will be displayed. This list will be the members of the category you chose. If you picked the correct category, the memorized goal word should be on the list. Simply find the goal word and press its button. The trial is now complete in terms of what you have to do. Think of the task as a search for the goal word selected by the computer. You have to search through the 2-level hierarchy shown in the diagram.

******************************************************

Please use the index finger of one hand on all trials. You decide which hand to use and then stay with it.

After each trial you will receive "correct" if your 2 responses identified the goal word correctly, and "error" if you make a mistake. On correct trials you will also receive your total response time in seconds.

******************************************************

Let's try an example:

The computer picks the word "football" and presents it on the screen:

FOOTBALL

The screen goes blank and then displays 6 category choices.
You then pick sports as the correct category for football by pressing the middle button on the right-hand side of the screen.

*********************************************************

Picking sports then gives you the complete list of sports displayed on the screen:

BASEBALL          TENNIS

GOLF              SWIMMING

FOOTBALL          SOCCER

Then select football by pressing the lower-left button.

*********************************************************
The word "correct" and total response time for the trial will then be displayed.

You will be performing 72 trials, split by a short break. Try to make your selections as quickly as possible, keeping errors to a minimum. The person with the lowest overall response times for today's task will receive a $5.00 bonus.

(If you understand these instructions, please push the upper left button and we will begin. If you have any questions, push the button on the desk.)

*******************************************************************************
The task you will be doing today and tomorrow is very similar to the one you did on the first day. In fact it is the same task, but involves different words and a different organization of the words. You will receive $3.00 for today's session. A bonus of $5.00 will be awarded to the person in each group who has the fastest overall time and the fewest errors over the next two sessions.

(Press upper left button to display next set of instructions.)

****************************************************************

Just as before, study the new word diagram marked "S" on the table top. You will have [2.8] minutes to study this diagram.

****************************************************************

Now that you have studied the diagram the task should be easy to follow. Just as before, you will be given a goal word chosen at random from the lowest level of the hierarchy shown in the diagram. You have to find the goal word by selecting the appropriate category descriptors in sequence until you arrive at the list containing the goal word.

****************************************************************
For instance, a correct trial would proceed as follows:

The computer picks the goal word:

SPAIN

You memorize the word and wait for the first menu.

The first menu gives you the choice of:

[TECHNOLOGY  GEOGRAPHY]

[BIOSCIENCE  SOCIOLGY]

You pick [geography] because it makes sense and because you know from the diagram that it leads down the structure to Spain.

The next choice is:

[EUROPE  LAND]

[ASIA  WATER]

You select [Europe].
The next menu you will see should have the goal word as one of the choices:

[SPAIN  POLAND  FRANCE  YUGOSLAVIA]

And you select Spain.

CORRECT  10.7

If you make a mistake somewhere in the sequence your final menu will not include the goal word. If you select the wrong goal word you will receive feedback stating such.

ERROR
If you make a wrong choice somewhere in the sequence and you realize it before you get to the final menu, press the "error" button on the desk. That will tell the computer that you know you made a mistake and will terminate the trial early. All error trials will be repeated once, later in the session. Errors are very costly in terms of your overall score for the session, so work as quickly as possible but avoid making mistakes.

*****************************************************************************

Please use the first finger of the same hand you used before for all responses. Using two hands may be tempting, but you will be consistently faster using only one hand.

*****************************************************************************

You will be doing several blocks of trials separated by rest periods. On occasion a few trials of a simple task will occur. You will receive specific instructions at that time.

If you are ready to begin, press the upper left button. If you have any questions press the button marked "error" and the experimenter will come and assist you.

*****************************************************************************
INSTRUCTION SET FOR THE 641 CONDITION

The task you will be doing today and tomorrow is very similar to the one you did on the first day. It is a search task, but involves different words and a different organization of the words. You will receive $3.00 for today's session. A bonus of $5.00 will be awarded to the person in each group who has the fastest overall time and the fewest errors over the next two sessions.

(Press upper left button to display next set of instructions.)

******************************************************************************

Just as before, study the new word diagram marked "S" on the table top. You will have 2.1 minutes to study this diagram.

******************************************************************************

Now that you have studied the diagram the task should be easy to follow. Just as before, you will be given a goal word chosen at random from the diagram. You have to find the goal word by searching through the entire word set presented on the screen. The words will be randomly located each time they are displayed. Your task is to locate the goal word and press the button associated with the block of 8 words containing it.

******************************************************************************
For instance, a correct trial would proceed as follows:

The computer picks the goal word:

**SPAIN**

You memorize the word and wait for the randomized word set.

******************************************************************************

You find the word Spain;

<table>
<thead>
<tr>
<th>ARISTOTLE</th>
<th>PLATO</th>
<th>CAESAR</th>
<th>CLEOPATRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>MACARTHUR</td>
<td>PATTON</td>
<td>ROOSEVELT</td>
<td>LINCOLN</td>
</tr>
<tr>
<td>CARVING</td>
<td>POTTERY</td>
<td>PAINTING</td>
<td>DRAWING</td>
</tr>
<tr>
<td>WALTZ</td>
<td>BALLET</td>
<td>SYMPHONY</td>
<td>OPERA</td>
</tr>
<tr>
<td>ERIE</td>
<td>HURON</td>
<td>PACIFIC</td>
<td>ATLANTIC</td>
</tr>
<tr>
<td>MATUREHORN</td>
<td>EVEREST</td>
<td>SAHARA</td>
<td>MOJAVE</td>
</tr>
<tr>
<td>CAMBODIA</td>
<td>INDIA</td>
<td>JAPAN</td>
<td>RUSSIA</td>
</tr>
<tr>
<td>FRANCE</td>
<td>MAGNETISM</td>
<td>YUGOSLAVIA</td>
<td>POLAND</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SNAKE</th>
<th>LIZARD</th>
<th>DOG</th>
<th>HORSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAILFISH</td>
<td>MARLIN</td>
<td>TROUT</td>
<td>SALMON</td>
</tr>
<tr>
<td>CANCER</td>
<td>STROKE</td>
<td>MUMPS</td>
<td>MEASLES</td>
</tr>
<tr>
<td>HEART</td>
<td>STOMACH</td>
<td>LIMBS</td>
<td>SPINE</td>
</tr>
<tr>
<td>ACID</td>
<td>WATER</td>
<td>CRYSTAL</td>
<td>METAL</td>
</tr>
<tr>
<td>NUCLEAR</td>
<td>FIRE</td>
<td>SPAIN</td>
<td>GRAVITY</td>
</tr>
<tr>
<td>CLOVER</td>
<td>GRASS</td>
<td>WHEAT</td>
<td>OATS</td>
</tr>
<tr>
<td>CABBAGE</td>
<td>LETTUCE</td>
<td>DAISY</td>
<td>ROSE</td>
</tr>
</tbody>
</table>

and press the third button from the left of the bottom of the screen.

******************************************************************************
The trial is now complete. You will receive feedback concerning the correctness and the time it took in seconds:

**CORRECT**

10.7

If you select the wrong block containing the goal word you will receive feedback stating such. If you don't find the word in 25 seconds, you will receive an error.

**ERROR**

If you forget what word you are searching for, press the "error" button on the desk. That will tell the computer that you know you made a mistake and will terminate the trial early. All error trials will be repeated once, later in the session. Errors are very costly in terms of your overall score for the session, so work as quickly as possible but avoid making mistakes.

Please use the first finger of the same hand you used before for all responses. Using two hands may be tempting, but you will be consistently faster using only one hand.

You will be doing several blocks of trials separated by rest periods. On occasion a few trials of a simple task will occur. You will receive specific instructions at that time.

If you are ready to begin, press the upper left button. If you have any questions press the button marked "error" and the experimenter will come and assist you.
The tasks you will be doing today and tomorrow are very similar to the one you did on the first day. In fact they are subtasks of the same task, that involve different words and a different organization of the words. You will receive $3.00 for today's session. A bonus of $5.00 will be awarded to the person in each group who has the fastest overall time and the fewest errors over the next two sessions.

(Press upper left button to display next set of instructions.)

************************************************************************

Just as before, study the new word diagram marked "S" on the table top. You will have [2.8] minutes to study this diagram.

************************************************************************

Now that you have studied the diagram the task should be easy to follow. Just as before, you will be given a goal word chosen at random from the lowest level of the hierarchy shown in the diagram. Given the goal word you will have two different tasks to perform.

1) Category match: On these trials you will be given a choice of [4] categories after the goal word is displayed. Your task is to choose the correct category for the given goal word. You make your choice by pressing the button next to the category. The trial is now complete. The choice of categories will be selected at random from the diagram, somewhere to the left of the goal words. There always will be one correct choice available on the screen.

************************************************************************
For instance, a correct category match trial would proceed as follows:

The computer picks the goal word:

**SPAIN**

You memorize the word and wait for the category menu.

********************************************************

The category menu gives you the choice of:

[TECHNOLOGY

GEOGRAPHY

BIOLOGY

SOCIOLOGY]

You pick geography because it makes sense and because you know from the diagram that it leads down the structure to Spain.

********************************************************

The trial is now complete. You will receive feedback concerning the correctness of your choice and the time it took in seconds:

**CORRECT**

2.5

********************************************************
If you make a mistake you will receive feedback stating such.

ERROR

********************************************************

If you forget the goal word, press the "error" button on the desk. That will tell the computer that you know you made a mistake and will terminate the trial early. All error trials will be repeated once, later in the session. Errors are very costly in terms of your overall score for the session, so work as quickly as possible but avoid making mistakes.

********************************************************

The second kind of task you will be doing having been given the goal word is:

2) Identity match: On these trials you will be given a choice of [4] goal words after the target goal word is displayed. Your task is to choose the target goal word from the list. Press the appropriate button and you will receive feedback similar to that of the category task.

********************************************************

You will receive several blocks of category match trials and identity match trials during your experimental session. Each block will be identified by a title in the middle of the screen, such as:

"IDENTITY MATCH TASK"

********************************************************
You will be doing several blocks of trials separated by rest periods. On occasion a few trials of a simple task will occur. You will receive specific instructions at that time.

If you are ready to begin, press the upper left button. If you have any questions, press the button marked "error" and the experimenter will come and assist you.

*****************************************************
INSTRUCTION PAGE FOR MOTOR CONTROL TASK FOR ALL CONDITIONS EXCEPT 641

Several times during each session you will be given a block of trials called the "location" task. After displaying the title "location task" a set of "XXX" strings will appear, one next to each response button. This is to point out all the potential locations of the stimulus. The screen will blank and only one "XXX" string will appear. Your task is to press the adjacent response button as soon as possible. The screen will blank again and another trial will begin. No feedback will be given. Trials will come in quick succession, so stay alert. Press the appropriate button as quickly as you can without making any errors.

(Press upper left button to begin.)
Several times during each session you will be given a block of trials called the "location" task. After displaying the title "location task" a set of "XXX" strings will appear in 8 blocks of 8, just like the word list. This is to point out all the potential locations of the stimulus. The screen will blank and only one "XXX" string will appear. Your task is to press the response button corresponding to the parent block as soon as possible. The screen will blank again and another trial will begin. No feedback will be given. Trials will come in quick succession, so stay alert. Press the appropriate button as quickly as you can without making any errors.

(Press upper left button to begin.)
LIST OF REFERENCES


Miller, G. A. The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 1956, 63, 81-97.


Nickerson, R. S. Response times with a memory-dependent decision task. *Journal of Experimental Psychology*, 1966, 72, 761-769.
