Finding Information on a Menu: 
Linking Menu Organization to the User's Goals

Brad Mehlenbacher  
Carnegie Mellon University

Thomas M. Duffy  
Indiana University

James Palmer  
Carnegie Mellon University

Abstract

Design paradigms often ignore the diverse goals users bring to the computer interface. Any human-computer interaction can be viewed as a marriage of two systems: The user begins the interaction by formulating an information goal, and the computer software meets that goal with a sometimes complex list of potential topic areas. The user then accesses that topic list through the computer interface. Part of the act of accessing the topic list is selecting a potential topic, and this action is often supported by a menu interface. Although research is pervasive on how best to organize menu items to facilitate learning, search speed, and reduced selection errors, little has been done to examine the impact of different types of user goals or cues on a menu's effectiveness. In a study using three distinct cues—direct match, synonym, and iconic—and two menu organizations—alphabetical and func-

Author's present addresses: Brad Mehlenbacher, Rhetoric Department, Carnegie Mellon University, Pittsburgh, PA 15213; Thomas M. Duffy, Learning Resources, Indiana University, Bloomington, IN 47405; and James Palmer, Apple Computer, Inc., 20525 Mariani Avenue, Cupertino, CA 95014.
tional—data suggest that (a) the functional menu is more effective than the alphabetical menu for the synonym and iconic cues, (b) learning occurs with both menu designs (i.e., selection speed increases rapidly across the five trial blocks), and (c) users make fewer errors with the functionally organized menu. The results, in general, encourage more rigorous investigation of the interaction between the tasks users bring to menu interfaces and the optimal design of those menus.

1. USER GOALS AND COMPUTER MENUS

Among the various methods designed to facilitate human-computer interaction, menu-driven interfaces have received an immense amount of use and attention over the last 10 years (Card, 1982; Giroux & Belleau, 1986; Robertson, McCracken, & Newell, 1981). The addition of menus to existing software, with the supposed guarantee of increased system friendliness, is a standard development in the computer industry; so the essential difference between, for example, CMS and PROFS, Dbase II and Dbase III, PageMaker 2.0 and PageMaker 3.0 (IBM, Ashton-Tate, and Aldus Products, respectively), is the addition or improvement of a menu. Their functionality is relatively unchanged, only the method of accessing the system's functions has been "enhanced."

The goal in the design of any menu should be to facilitate the user's ability to make a choice quickly and accurately. Human-computer research has tended to focus on two primary issues as most relevant to achieving this goal: the hierarchical organization of menus and the organization of items on any one menu. The major emphasis for hierarchical organization is on the proper
balance of menu breadth (the number of items one must scan on a given menu) and menu depth (the number of menus one has to pass through to get to the desired information) because they affect search time and selection accuracy.

There are three variables most relevant to determining the proper balance of depth and breadth: the number of items an individual must scan (search time), the number of choices the user must make (decision and motor response time), and system response time to those choices (system time). Adding more menus reduces the number of items a person must scan, but it increases the number of user and system responses that have to be made. For example, with 64 items all on one menu, the user, on the average, will scan 32 (50% of the) items before arriving at the correct choice (Lee & MacGregor, 1985), but there will only be one user and one system response. Alternatively, we could create a hierarchical structure of the two item menus so that the user makes a series of six binary choices (six user and six system responses), to get to the correct selection. With the six choices, however, the user will (on the average) scan only 1 item per menu for a total of 6 items scanned to get the correct selection. Calculations of the optimum balance of breadth and depth, based on an assumption of self-terminating search (i.e., the user quits the search of the menu when the item is located), indicate that the optimum number of items per menu is between 4 and 8 for a wide variety of stimulus sets (Lee & MacGregor, 1985). In general, research has supported those calculations (Kiger, 1984; Miller, 1981).

Concern with the breadth–depth issue has tended to dominate the research literature (Allen, 1983; Kiger, 1984; Landauer & Nachbar, 1985). However, Paap and Roske-Hofstrand (1986) suggested that the organization of items on a given menu may be the more important variable for optimizing speed and accuracy of selection. The research on the breadth–depth tradeoff has been based, for the most part, on the assumption that for any one menu, the user begins at the beginning and progresses through it until the needed item is located. That is, the research presumes that the items on the menu are not organized in any meaningful fashion so as to support a systematic search strategy. However, if the items on a menu are organized in some way and if the user understands or learns the structure of that organization, then the user can skip over groups of alternatives and visually isolate the approximate location of the target item. Given an effective organization, the assumption in the breadth–depth tradeoff studies that the user will (on the average) scan a minimum of half the items is invalid. With a well-defined organization, there need not be a difference between a deep menu structure (a few items on many menus) and a broad menu structure (many items on a few or even a single menu) in terms of the number of items scanned. Indeed, the deeper structure could even require the user to scan more items than would be required on a single, well-structured menu. Thus, the organization of the items on the
menu, rather than the depth of the menus, becomes the key determinant of the efficiency of a search. Certainly, research has shown that the arrangement of menu items affects the amount of time users spend searching for items (Allen, 1983), the number of correct selections made (Lee, MacGregor, Lam, & Chao, 1986), how easily the menu can be learned (Card, 1982), and the overall user satisfaction with the product (Magers, 1983). The issue, however, is identifying the most effective organization of the items for the population of users. An operational definition of most effective is that the organization minimizes search time and selection errors. From a design viewpoint, this has two implications. First, the number of groups and the number of items per group must be optimized so as to reduce the number of items a user must search in selecting an answer. This optimization will occur when the number of groups equals the number of items per group; that is, the number of groups equals the square root of the total number of items on the menu.

The second issue, and the issue researched and reported here is the basis on which the items are grouped. Grouping alone may be useful, because spacing between groups is a landmark that can be used in maintaining one's place during a search (Parkinson, Sisson, & Snowberry, 1985; Rubens & Krull, 1985). However, the main benefit of grouping will come when the individual can use the structure to guide the search. There are two basic strategies for organizing items on a menu: alphabetical grouping and grouping based on semantic or functional relationships between the menu items.\footnote{Actually, the functional grouping may be based on an analysis of the functions in the application program, on the individual's use of those commands, or on the combination of user and application considerations. Grouping of items based on the physical layout of information on the computer console is an example of a grouping based primarily on the application, whereas a grouping based on frequency of use is an organization based primarily on the user. Our reference to functional grouping in this article is grouping based on combined user-application considerations, for example, when commands are grouped based on the semantic characteristics of the user's tasks.} Research has consistently indicated that either of these organizational strategies is better than no systematic arrangement (Card, 1982; McDonald & Schvaneveldt, 1988; McDonald, Stone, & Liebelt, 1983). However, there have been conflicting results when the two approaches have been compared.

For example, Parkinson et al. (1985) failed to find any difference in accuracy or speed when comparing the performance of alphabetically to functionally organized menus. In the Parkinson et al. (1985) study, 64 items from four naturally occurring groups were presented on a single screen. The organization was manipulated within the four groups of items. The dominant overall organization into four semantic categories may have negated any differential effect of within category differences. However, McDonald et al. (1983) found an effect of subgroup organization using a similar stimulus set.

In contrast to these findings, Card (1982) found that selection was faster
with an alphabetical arrangement than with a functional arrangement. This study employed a menu of 18 word processing terms arranged in a single, vertical menu. Subjects were given a command name and asked to select the appropriate target item from the menu. Card's (1982) results have led some authors of interface design texts to include a guideline stating that menus accessing simple reference documentation should be organized alphabetically (Brockmann, 1986, p. 230) and have, therefore, added credence to the traditional, alphabetical method of organizing menu items.

Finally, in contrast to all of the results presented thus far, MacDonald et al. (1983) found that semantic or functional ordering is superior to alphabetical for searching and selecting menu items. McDonald et al. (1983) presented a 64-item menu with items from naturally occurring categories. Subjects were given either a word on the menu (direct match) or a definition of one of the words. Regardless of the stimulus, however, the functional grouping led to the fastest and most accurate selection. Consistent with the Card (1982) findings, this result has also impacted the presentation of development guidelines, except that now we find guidelines recommending a functional rather than an alphabetical organization (see, e.g., Shneiderman, 1986, p. 111).

We argue that the design of a menu must be based on a consideration of the tasks the user must perform in making a menu selection. An analysis of these tasks may help explain the conflicting results and lead to the development of more ecologically valid guidelines. There are two cognitive tasks in making a menu selection. First, the user must represent what information is required. That is, the user must represent the goal in searching the menu. Second, the menu options must be reviewed to determine which one best meets the information needs of the user.

The user's representation of the information need will depend on his or her knowledge base. Here we find it useful to employ Shneiderman's (1986) distinction between semantic and syntactic knowledge (for a more detailed discussion, see Duffy, Mehlenbacher, & Palmer, 1989). Syntactic knowledge is knowledge of the specific requirements of a particular system; that is, knowing the command names specific to the particular application. Semantic knowledge is broken into two categories: computer concepts and task concepts. Task concepts are those concepts conceptualized apart from the computer (e.g., knowing the writing, editing, and organizational tasks involved in preparing a proposal). Knowledge of computer concepts involves knowledge of the way a task is generally structured in a computer program (e.g., the kinds of word processing and file manipulation tasks that will be involved in preparing the proposal on a computer). It is important to recognize that the user's understanding of the task based on computer concepts could be very different from the understanding of the task based on task concepts (Breuker & de Greef, 1985).
With this distinction in mind, we can now describe three ways in which the user may represent the information he or she is looking for on a menu. First, the user may have syntactic knowledge and, therefore, know the exact menu item he or she desires. This would be characteristic of most task representations by expert users of the particular application software. Given this goal representation, the user's second task, searching the menu, consists of finding an exact match for the goal representation. Here the perceptual demands are much greater than the cognitive demands of the search.

The second type of representation is based on knowledge of the computer concepts. This is characteristic of a transfer user or an occasional user. That is, the user understands how the task is generally broken down in computer programs and knows the kinds of names attached to the tasks. However, the particular naming conventions in this program are unknown. For example, the user may realize that moving a paragraph generally involves cutting and pasting, but he or she does not know whether the relevant commands are "set mark," "cut," et cetera. The cognitive demands of the menu search task are greatly increased, given this problematic representation. The user must now review each menu item and judge whether or not the command is a synonym for the representation.

Finally, the user may have only acquired task knowledge and not a representation of the task in terms of computer concepts. He or she knows what needs to be accomplished but does not know how to represent the task in computer terms. This type of representation is common to the novice computer user and may even be common to a user who has some computer experience but who is unfamiliar with a particular class of computer application programs.

In summary, the user may approach the menu looking for a direct match, looking for a synonym for the computer concept, or looking for some computer concepts that seem relevant to the task. If the menu designer is to support the user's search, then the design must accommodate the user's problem representation and the knowledge base the user is bringing to the search task. If the user is experienced in the domain and is looking for a direct match for the stimulus in his or her goal representation, then an alphabetical organization would typically be most effective. That is, an organization based on the alphabetical sequence is extremely well-learned, and the user should be able to jump rapidly to the correct location on the menu. An organization based on semantic relationships between items may be just as effective as the alphabetical organization if it is as well-established; it may even be more effective if the organizational structure is even more firmly established.

These considerations may help resolve the conflicting results obtained by Card (1982) and McDonald et al. (1983). Both studies asked users to find a direct match on the menu. Card (1982) found the alphabetical organization most effective whereas McDonald et al. (1983) found the functional organi-
zation to be most effective. The distinguishing feature is that although Card (1982) employed word processing concepts, McDonald et al. (1983) used naturally occurring categories of terms. McDonald et al. (1983) used a card sort strategy to ensure that the organization between categories was maximally distinctive from the relations within categories for the users.

Because the alphabetical structure has been learned and used since childhood in a wide variety of contexts, it is unlikely that any semantic organization of computer concepts will provide as strong a structure for guiding the search for a particular, known term. Thus, we expect that the alphabetical organization provides a more usable structure in the Card (1982) study. However, the semantic relations used by McDonald et al. (1983) were learned since childhood and are used in even a wider variety of circumstances than the alphabetical sequence. It is reasonable to assume, therefore, that the semantic organization could provide a more supportive structure than the alphabetical organization—or at least that results would be inconsistent when two well-learned structures are compared.

Most research has focused on the direct match search task. However, McDonald et al. (1983) also used the definition of words in the menu as the cue. Thus semantic, not just syntactic, knowledge was required. As we predicted, the functional organization was even more effective, relative to the alphabetical organization, in supporting this task.

The focus of our research is on the design of menus for computer systems, that is, for menus in which the semantics are not overlearned. McDonald et al. (1983) found that for videotex-based menus, a semantic organization was most effective regardless of the user's task. We predict that, with computer menus, the most effective menu organization depends on the user's task or problem representation. For a direct match, characteristic of the experienced user, an alphabetical organization should be most effective. However, when the user only has semantic knowledge, a functional or semantic organization will be most effective provided, of course, that it reasonably reflects the user's semantic organization.

Our research compared the effectiveness of two types of command-based menu organizations in supporting the efficient location of the appropriate command given one of three kinds of cues aimed at simulating the different types of problem representations that users bring to a menu interface. The three cues were an exact command to be located on the menu, a synonym for the menu item, or an iconic representation of the action performed by the command. We used a menu containing word processing and formatting commands and tested users who had some computer and word processing experience. The users, therefore, were equivalent to transfer users with a low level of general experience. The iconic representation involved the presentation of before and after texts in which the before showed graphically (without words) the kinds of changes to be made and the after showed the text with the
Figure 1. The alphabetically organized menu.

<table>
<thead>
<tr>
<th>MENU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bold</td>
</tr>
<tr>
<td>Cancel</td>
</tr>
<tr>
<td>Center</td>
</tr>
<tr>
<td>Copy</td>
</tr>
<tr>
<td>Delete</td>
</tr>
<tr>
<td>Exit</td>
</tr>
<tr>
<td>Font</td>
</tr>
<tr>
<td>Find</td>
</tr>
</tbody>
</table>

Click the mouse over the correct menu item.

changes completed. We expect that the users, given their level of experience, have greater difficulty matching a menu item to the iconic stimulus than to the synonym.

2. EXPERIMENTAL METHOD

2.1. Design and Materials

Three experimental variables were manipulated in a $2 \times 3 \times 5$ mixed factorial design. The first between subjects variable, menu organization, consisted of an alphabetically organized menu (see Figure 1) and a functionally organized menu (see Figure 2). Both menus had 22 items separated into three columns. The number of menu items approximates many of the conventional word processing menus with between 15 and 25 items to choose from (see, e.g., WordStar, MicroSoft Word, and EZ). In the functional menu, a category heading in bold type spanned each group of items.

The second between subjects variable, cue type, consisted of three types of cues meant to capture the different types of goal tasks the users bring to a word processing menu. The three types of cues were a direct match cue, a synonym cue, and an iconic cue. A direct match cue is, "I know the word that the application uses and I want to find it." For the direct match cue, the subject is shown, for example, the word *center* and asked to select the command "Center" on the menu. A synonym cue is, "I know what I want to do, but do not know the exact command that the application uses." For the synonym cue, the subject is given the phrase "Move text to middle" for the desired correct response of selecting the command "Center" on the menu. An
Figure 2. The functionally organized menu.

<table>
<thead>
<tr>
<th>Changing Text</th>
<th>Formatting</th>
<th>Control Commands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insert</td>
<td>Bold</td>
<td>Windows</td>
</tr>
<tr>
<td>Delete</td>
<td>Italic</td>
<td>Undo</td>
</tr>
<tr>
<td>Replace</td>
<td>Underline</td>
<td>Cancel</td>
</tr>
<tr>
<td>Move</td>
<td>Upper</td>
<td>Print</td>
</tr>
<tr>
<td>Copy</td>
<td>Font</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Finding Text</th>
<th>Alignment</th>
<th>Files</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find</td>
<td>Center</td>
<td>Get</td>
</tr>
<tr>
<td>Jump</td>
<td>Margin</td>
<td>Put</td>
</tr>
<tr>
<td></td>
<td>Justify</td>
<td>Exit</td>
</tr>
</tbody>
</table>

Click the mouse over the correct menu item.

iconic cue is, "I have a picture of what I want the document to look like, before and after, and I want to find the command that will allow me to accomplish it." In the iconic cue, the subject is shown a picture of a text before and after a change had been made to it (see Figure 3).

The within subjects factor was Trial Block which consisted of five blocks of 14 trials for each experimental condition.

There were also three dependent measures: study time, selection time, and accuracy of the selection. Study time is the amount of time the subject spent studying the cue before requesting the menu. The subjects were told they could study the cue as long as they wanted, but once they requested the menu they should try to respond as quickly as possible. Given these instructions and the separation of the two intervals, we assume that the study time is a reasonable index of the difficulty of problem formulation.

Selection time is the time from when the menu is presented to the time the subject made a selection. This time is assumed to reflect the variety of cognitive processes involved in trying to find a menu item that will satisfy the need as formulated: the reformulation of the problem if no match is found, perceptual search time, motor response time, and decision making.

2.2. Subjects

Seventy-five university students acted as subjects for the experiment. Subjects were screened beforehand, and only individuals who had some
Figure 3. Iconic cue representing Center.

<table>
<thead>
<tr>
<th>BEFORE</th>
<th>AFTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>The ability to remember information is an essential and vital ingredient of human information-processing. Even the simplest adaptive system cannot function without memory, and people are complex adaptive systems.</td>
<td>The ability to remember information is an essential and vital ingredient of human information-processing. Even the simplest adaptive system cannot function without memory, and people are complex adaptive systems.</td>
</tr>
<tr>
<td>The task of designing systems that do not overload memory is a difficult one. In this section, we discuss three memory subsystems called sensory memory, working or short-term memory, and long-term memory.</td>
<td>The task of designing systems that do not overload memory is a difficult one. In this section, we discuss three memory subsystems called sensory memory, working or short-term memory, and long-term memory.</td>
</tr>
</tbody>
</table>

computer word processing experience were permitted to participate. Before they began the experiment, they were asked to fill out a short questionnaire asking for basic demographic information (year in college, major or program name, etc.) and word processing background information (years of experience, how many times they had used a word processor in the last 2 weeks, and which word processor or processors had they used). Subjects were paid $2 per session or received course credit for their participation.

Student responses to the preliminary questionnaire indicated that their experience with word processors ranged from 6 months to 5 years ($M = 2.3$, $SD = 1.3$). The types of word processors used by the subjects varied and included AppleWriter, Ed, EMACS, Final Word, MacWrite, MicroSoft Word, WordStar, and Xedit. The average number of reported uses during the last 2 weeks was 5.8 ($SD = 4.8$). An analysis of previous experience and reported word processing over the last 2 weeks showed that there were no significant differences across the six experimental conditions.

2.3. Procedure

Eight groups of subjects were tested with 5 to 12 subjects per group. The subjects in a group were randomly assigned to conditions with the restriction that all six conditions were represented about equally. All instructions and test material were presented on a microcomputer.

Subjects were told that we were evaluating alternative designs for word processing menus. Then they placed their diskette in the machine and
individually progressed through the condition-specific instructions and testing. They were told that they would be shown a word processing term (or a text with editing marks and then that same text with the editing completed). Their task was to find the word on a menu that was the same as (or that described) the cue. The instructions stressed the importance of selecting the correct term on a menu as well as making selections as quickly as possible when the menu appeared.

The subjects received five practice trials using animal names (and pictures for the iconic condition) after reading the instructions. The trials in practice and in the main experiment progressed as follows. A ready signal was presented for 2 s and was then replaced with the cue. The subject studied the cue and then clicked the mouse in a box at the bottom of the screen when he or she was ready to see the menu. When the menu appeared, the subject's task was to click on the correct menu item for the cue. No feedback on the correctness of the choice was provided. After the selection, the menu was replaced with the ready signal to mark the beginning of the next trial.

The word processing menu was presented for 1 min at the beginning of testing, and the subjects were instructed to familiarize themselves with the terminology and format. There were five blocks of 14 trials in testing. The same 14 commands were tested in each block of trials and for all conditions. The subset of 14 commands were spread across each of the three columns in the alphabetical menu and each of the six functional groups in the functional menu. The commands, which were randomly ordered for each of the five trial blocks, were: Bold, Center, Copy, Delete, Exit, Find, Insert, Italic, Justify, Move, Print, Underline, Upper, and Windows (see Figures 1 and 2).

3. RESULTS AND DISCUSSION

3.1. Problem Formulation

A $2 \times 3 \times 5$ mixed factorial, unweighted means analysis of variance (ANOVA) performed on the mean study time yielded significant main effects of cue, $F(2, 68) = 14.4, p < .001$, and trial blocks, $F(4, 272) = 112.1, p < .001$, as well as a significant cue $\times$ trial block interaction, $F(8, 272) = 44.13, p < .03$. These results were as anticipated. Because the same cues were presented in each block of trials, we expect study time to decrease across trials as illustrated in Figure 4. Furthermore, the direct match cue require the least user "transformation" of the cue into a potential menu item and, hence, less study time, whereas the iconic cue requires the most user transformation and, hence, the most study time. The data in Figure 4 are consistent with this expectation. Finally, with practice, each cue should eventually elicit a menu item response quite directly and thus, as confirmed by the interaction, the cue effect should dissipate across trials.
There was also an unanticipated main effect of menu, $F(1, 68) = 3.97, p < .05$, as well as a significant three way interaction, $F(8, 272) = 2.29, p < .03$. The main effect of menu organization was due to a somewhat fast study time in the functional menu condition. Although the difference persisted across trial blocks, the primary difference was in the first trial block where the mean study times were 6.27 s ($SD = 4.56$) for the three alphabetical menu conditions and 5.19 ($SD = 3.29$) for the three functional menu conditions. Furthermore, the advantage of the functional menu was greatest for the iconic cue conditions ($M = 11.33$ and $8.19$ s in the first trial block for the alphabetical and functional menus, respectively) but was very minor with the direct match cue (a difference of .09 s in the first trial block). The significant three-way interaction reflects this effect.

There are two possible explanations why the functional organization facilitates the problem formulation phase. First, the functional organization provides a structure that facilitates the user's ability to learn the command names on the menu. Items in semantic categories are easier to learn than uncategorized items. As a result of this learning, the user should possess a greater repertoire of commands on which to base his or her information needs and ultimate decision. Second, the functional menu provides conceptual bins or chunks that help the user classify the cue. These categories may be easily learned and thus be used to allow the user to formulate the problem at this more general level and move to the menu to evaluate specific alternatives.
Both of these interpretations are consistent with a general model that argues that functionally organized menus provide a schema for thinking about and learning menu items. Both interpretations also predict (as was found) that the functional menu would have a significant effect on the behavior of the user who studied the iconic cues and virtually no effect on the user who studied the direct match cues.

3.2. Menu Selection Time

Once the participant formulated the problem, he or she requested the menu and made a menu selection as rapidly as possible. We submitted the menu selection time to the $2 \times 3 \times 5$ mixed factorial, unweighted means ANOVA in which the between subjects variables were menu organization and cue and the within subjects factor was Trial Block. The only main effect to achieve significance was Trial Block, $F(4, 272) = 49.6$, $p < .001$, and this simply reflected the fact that the performance of all the groups improved across trial blocks. The effects of menu organization and cue type appear in the interaction with trial blocks. The Menu $\times$ Trial Blocks interaction yielded an $F(4, 272) = 2.51$, $p < .05$, whereas the Cue $\times$ Trial Blocks interaction yielded an $F(8, 272) = 4.39$, $p < .001$. The three-way interaction also reached significance, $F(8, 272) = 2.3$, $p < .02$.

The interaction of menu organization and trial blocks is illustrated in Figures 5, 6, and 7 for the direct match, synonym, and iconic cues, respectively. As can be seen in the figures, the groups all come together in the later trial blocks, with mean performance ranging between 2.44 s and 3.08 s on the last trial block. A post hoc comparison using Duncan's Multiple Range Test failed to yield any significant differences between the six groups on this last trial block. The primary effects are seen in the early trial blocks.

Looking at the first trial block, we see that the alphabetic menu leads to faster selection time than the functional menu only under the direct match condition, that is, when the cue is an item from the menu. As we expected, the participants are well-versed in the alphabetical structure and the ability to locate a command based on an alphabetical organization transfers easily to this task. If the functional structure of word processing terms was also well-known we expect the functional organization to be equally effective.

The opposite effect occurs when we look at the synonym and iconic conditions. In both cases, the functional organization leads to faster menu selection time. The difference between the functional and alphabetical menus is very large for the iconic condition (a difference of almost 3 s in the first trial block), and the effect persists across all blocks.

In concluding the presentation of the time results, note that the data for the total time (study plus menu selection times) yields relations almost identical to those illustrated in Figures 5, 6, and 7.
3.3. Selection Errors

The participants in the direct match conditions made virtually no errors in selecting menu items ($M = 0.04$ across the five trials). Because they were simply selecting the exact item presented, no errors were expected. However, participants in the synonym and iconic conditions had to select the menu item they felt would accomplish the task represented or that meant the same as the synonym presented. There were errors in these conditions and, as expected, there were more errors with the iconic cues than with the synonyms.

Of interest is whether or not the functional organization led to fewer selection errors. We anticipate that the shared meaning achieved by a functional grouping helps the participant infer the meaning of a command and, therefore, make fewer errors attempting to link the task (represented by the cue) with the appropriate command.

Mean errors per Trial Block are presented in Figure 8 for the synonym and iconic cue conditions. As can be seen, the functional menu leads to a lower error rate with both cue types. In addition, although error rate decreases across trial blocks, participants still made fewer errors with the functional menu even on the last trial block.

The error rate for the synonym condition was too low and had too many 0 scores to submit to an ANOVA. However, the data from the iconic cue condition was analyzed in a $2 \times 5$ mixed factorial weighted means ANOVA
Figure 6. Mean menu selection time for the alphabetical and the functional menus in the synonym cue condition across the five trial blocks.

Figure 7. Mean menu selection time for the alphabetical and the functional menus in the iconic cue condition across the five trial blocks.
Figure 8. Mean number of selection errors for the alphabetical and the functional menus in the synonym and iconic cue conditions across the five trial blocks.

![Graph showing mean number of errors per trial block for different conditions.]

with Menu Organization and Trial Blocks as the factors. Trial Blocks yielded a significant effect, $F(4, 92) = 8.92, p < .001$. Similarly, the main effect of Menu Organization was significant, $F(1, 23) = 4.78, p < .05$. The interaction of Menu Organization with Trail Blocks, however, did not reach significance. The results confirm the relations apparent in Figure 8.

Because the subjects did not receive feedback on their responses, we did not anticipate large learning effects across trial blocks. Indeed, as can be seen in Figure 8, the trials effect is primarily due to an improvement in performance from the first to the second trial block. Thus, the improvement seems to be due to the subjects learning about the full stimulus set during the first trial block and being able to use that information to correct some of their responses during the second trial block.

It is possible that the difference in errors between the cue and menu conditions may account for the difference in selection time. That is, it may be that the participants take longer when making error responses and, because they made more errors with the alphabetical menu, mean selection time is also longer. We assumed that the functional organization facilitates search: The user can look at a category title or an exemplar of the category and quickly judge whether or not it is a relevant group to search. However, if the selection time effect is restricted to error responses, we must reject this hypothesis and conclude that the effect of the functional organization is due to the error reduction that occurs through specific semantic effects; by combining correct
**Figure 9.** Mean time for selection without errors for the alphabetical and the functional menus in the iconic cue condition across the five trial blocks.

and incorrect selection times, we may be looking at decision making rather than actual search time.

To test this hypothesis, we reanalyzed the time required to make the menu selection, but this time, we only included the time data for correct selections. Because the majority of errors were in the iconic condition, we only conducted this reanalysis for the iconic cue conditions. A 2 × 5 unweighted means ANOVA of the mean time per trial block for correct choices yielded a significant main effect for Trial Blocks, F(4, 92) = 25.02, p < .001, and a significant Menu × Trial Block interaction, F(4, 92) = 3.28, p < .05. The results illustrated in Figure 9 parallel the time data for all responses, confirming that the time effects are not due to longer times on error responses.

4. CONCLUSION

4.1. Implications for Interface Design

The results confirm our hypothesis that the most effective organization of a menu depends on the user's knowledge base and information goal. For the individual who knows what command he or she wants, the alphabetic structure will typically be most effective. If the alphabetical structure of information is the strongest cognitive organization of the set of commands, then it will be most effective. This was the case for our menu, and we expect
it to be the case for most menus from application programs. If the user “knows” the knowledge domain very well (i.e., has a long history of experience with the item set), then we may expect his or her functional model of the organization of items in the domain will be equivalent to an alphabetical organization. Hence, alphabetical and functional organizations would be equally effective in a direct match task (presuming of course that the appropriate functional organization is presented; Palmer et al., 1988). This could be the case for the well-practiced expert using an application program. It would also be the case for any user searching a menu in a retrieval system dealing with common, well-learned information (see, e.g., Hollands & Merkle, 1987).

More typically, however, users will not be well-practiced nor will they have an overlearned organization of the knowledge domain (as is found with popular databases). For this more typical case, a functional organization is more effective when the user does not know exactly what command is required—as is often the case with novice and intermediate users. Indeed, our results suggest that the greater the user's uncertainty about the desired command, the more effective the functional organization will be. In essence, we suggest the following: (a) that menu organization will make little difference for the expert (who knows the command he or she wants and has a very strong functional organization of the knowledge domain), (b) that an alphabetic organization will probably be most effective for a wider range of experts (who generally know the command they want), and (c) that a functional organization will be most effective for novices (who do not know what command they need). However, casting this as an expertise issue is really a means of summarizing the more precise characterization in terms of user knowledge and task. Therefore, the relation to expertise just outlined is made with an important proviso.

We recognize that user expertise refers to the particular command set: Expertise is not a general characteristic, even in terms of particular application software. For example, even “experts” of complex word processors like Xedit or EMACS typically do not use a large set of the program’s commands. There are subsets of EMACS commands or functions with which the EMACS expert may well be a novice. As long as the expert restricts the use of an application like EMACS to the familiar commands or functions, which is unlikely because complex systems tend to invite complex learning situations, the alphabetical organization should be as effective as the functional organization. However, if these experts begin to extend their use of Xedit or EMACS still further and begin using new commands, then we predict a functional menu is most effective. In fact, because expertise involves extensive learning, we tend to argue that by the time novices and intermediate users reach such a level, the type of menu organization might as well have been functional all along.
4.2. Limitations

The cue conditions were meant to represent the problem formulation task of users with different levels of experience with word processing programs. In this sense, the iconic cue was designed to represent individuals who have never worked with word processors and simply think of the task in terms of the change they want to make. Of course, it is assumed that this is the novice’s starting point and that he or she will label the image—however, the label will tend to vary more and be more potentially off track than will the label generated by the transfer user (the synonym condition). Although we feel that the experimental conditions reasonably represented the tasks of these types of users, it is a simulation and must be treated as such when considering generalizations to real user contexts.

In particular, subjects may have had difficulty interpreting the iconic stimuli simply because of our skills, or lack thereof, in designing these stimuli. Thus the cues may have been difficult for more than one reason. Although this may have increased the difficulty of the iconic cues overall, we do not see any reason to suspect that the interaction between cue type and menu organization would be affected (i.e., the effect is simply one of increasing the overall difficulty of the iconic stimuli regardless of menu type).

Another experimental component, the use of trials, must also be considered in generalizing the results. Performance of our groups comes together quite rapidly such that differences disappear after just a few trial blocks. As a consequence, many of our hypothesized effects due to menu, cue, and the interaction of menu and cue only occurred in interaction with trials rather than as overall effects. The conclusion one might draw, then, is that all of the effects are short-lived and not to be worried about in the real world.

We pose the opposite argument; the effects on the early trials are relevant to a wide range of real world contexts. That is, the experimental context provides a very compressed experience within a very limited context. Our users worked continuously at the task, selecting one stimulus after another. This is quite different from the normal work experience in which many commands do not get used at all during a 1-hr work effort and wherein software programs are only used once a week. Also, in the normal work environment the cues and context for a particular command are continually changing. We used consistent stimuli from trial to trial, but of course the user of computer software is always facing new stimuli and new problems. Expertise simply does not develop as rapidly in the real world as is suggested by the increase in performance across trial blocks.

Finally, as is characteristic of most studies in the field, we worry about the generality of our findings to alternative sets of materials and alternative designs of that material. When one compares single instances of a generic
category—be it the menu item set as an instance of menus, the functional organization as an instance of functional, or the particular set of cues as instances of a user's initial representation—it is not possible to know if the differences are due to the instances or to the generic category. There is another side to this issue that gives us even greater concern: the clarity of our understanding of the general category. In this particular case, we are truly concerned about the notion of functional menus, lest any nonalphabetical grouping be called functional and hence be expected to facilitate user performance. Although we did not address the issue in this article, the design of effective functional menus requires the careful analysis of the audience, the tasks they will perform with the software, and their interpretation of the terms in the menu (Palmer et al., 1988).

Support. This work was funded in part by the U.S. Army Human Engineering Laboratory, Aberdeen Proving Grounds, MD, under contract DAAA1-86-K-0019; the contract monitor was Maureen Larkin. The views expressed in this article are ours and do not necessarily reflect the views of the U.S. Army.

REFERENCES


---


—Editor