

Memory for hierarchical menus: Effects of study mode

KENT L. NORMAN

University of Maryland, College Park, Maryland

and

JEFFREY P. SCHWARTZ

Grumman-CTEC, Inc., McLean, Virginia

Computer users need training on menu selection systems to remember how to access target functions. This experiment investigated the effects of studying documentation before searching for target words in a content-free menu. Subjects studied one of the following: sequences of choices to arrive at a desired target (command-sequence mode), individual display screens (menu-frame mode), a diagram of the menu tree (global-tree mode), or the menu system by actually selecting alternatives (trial-and-error mode). Although the global-tree and command-sequence groups found the most target words during the test phase, the four groups did not differ significantly. On the other hand, differences in the ability to recall menu terms were significant, with the command-sequence and menu-frame groups recalling the most terms. Results indicated that the type of training fundamentally affected the type of information used to find targets.

Menu selection systems are being used increasingly to facilitate the use of computers by novice users (Shneiderman, 1980). No matter how well written a menu selection system may be, however, novices can never be sure of what an alternative will do or where it will lead until they have selected it. Learning is an inherent part of using a menu selection system, no matter how self-explanatory the system may be.

Several researchers examined conditions that facilitate users' learning and performance on menu structures. Highly discriminable objects yield up to 95% correct choices in menu-based search (e.g., Allen, 1980; Miller, 1981). For real systems, the value drops to around 80% (e.g., McEwen, 1981), and this level is typically maintained only for choices that lead to goals that are, at most, three levels deep. Training methods have also been studied. Billingsley (1982) compared the performance of subjects who learned menu structures through a map of the entire menu, individual choice sequences, or trial and error. Choice speed and accuracy of the map group improved over time, whereas performance by the sequence group declined across trials. Performance of the trial-and-error group was the worst of the three groups. This result is especially relevant for teaching menu structures to new and casual systems users, since their learning is typically accomplished via trial-and-error exposure.

The experiment reported here investigated several modes of learning a menu selection system. The goal for the user was to learn the sequence of choices to get to a target node in the menu tree. The hypothesis was that mode of study would affect search performance and, more important, that mode of study would affect recall and reconstruction of the menu hierarchy. Four learning procedures were used in which subjects studied "documentation" on the menu system. In the *command-sequence mode*, subjects studied the sequence of correct choices that led to each target node. In the *menu-frame mode*, subjects studied the menu frames listing sets of items. In the *global-tree mode*, subjects were exposed to a tree diagram showing all menus and target nodes. Finally, in the *trial-and-error mode*, subjects were told to study by searching through the menu system itself to find out how to reach selected targets. A content-free menu system was used in which there were no meaningful relations between terms from one level to another. Consequently, any differences among study modes cannot be attributed to prior knowledge about the menu structure.

METHOD

Subjects

Subjects were 60 students in an introductory psychology course at the University of Maryland. Fifteen subjects were randomly assigned to each of the four conditions.

Materials

Menu frames were composed of five-letter paralogues for the main Level 1 menu, three-letter trigrams for Level 2 menus, and high-frequency common nouns for Level 3 target words. The paralogues were of low meaningfulness according to Taylor and Kimble (1967). The tri-

This research was supported by a grant from Control Data Corporation to the Human/Computer Interaction Laboratory in the Center for Automation Research, University of Maryland. We gratefully acknowledge Susan Haugh and Roberto Cofino for their assistance in data collection. Requests for reprints should be sent to Kent L. Norman, Department of Psychology, University of Maryland, College Park, MD 20742.

grams were CVCs within the 70-78 range on the Krueger scale (see Underwood & Schulz, 1960, pp. 309-317) and within the 31-39 range on the Archer (1960) scale. Target words were "A" and "AA" nouns according to Thorndike and Lorge's (1944) word count. The menu tree had three unnumbered alternatives at each of three levels. The structure of the menu tree is shown in Figure 1. Each of the 13 menu frames was printed with 1/2-in. letters on a 3 x 5 in. index card. Target words were also printed on cards. For the command-sequence mode, each of the 27 command sequences was composed of a paralog, a trigram, and a noun. Each set of three was printed on a separate card. The menu-frame mode used the 13 menu frame cards. Finally, for the global-tree mode, the menu structure shown in Figure 1 was printed on a poster board using 1/2-in. letters.

Procedure

The menu task was explained to the subjects, and they were allowed 5 min to study. In the command-sequence mode, subjects studied the 27 command-sequence cards. In the menu-frame mode, subjects studied the 13 menu-frame cards. In the global-tree mode, subjects studied the menu tree. Finally, in the trial-and-error mode, subjects were shown the main menu, they selected an alternative, and then they were shown the next menu with target words. They repeated this selection process for the entirety of the study period.

In the test phase subjects were shown a target word and the start frame. They selected an alternative by giving its number and were then shown the next frame. Again, they selected an alternative and were then shown the terminal node. At this point they could determine whether or not they had found the target word. Subjects were tested on all 27 target words in a random order. Study materials were not available during the test phase.

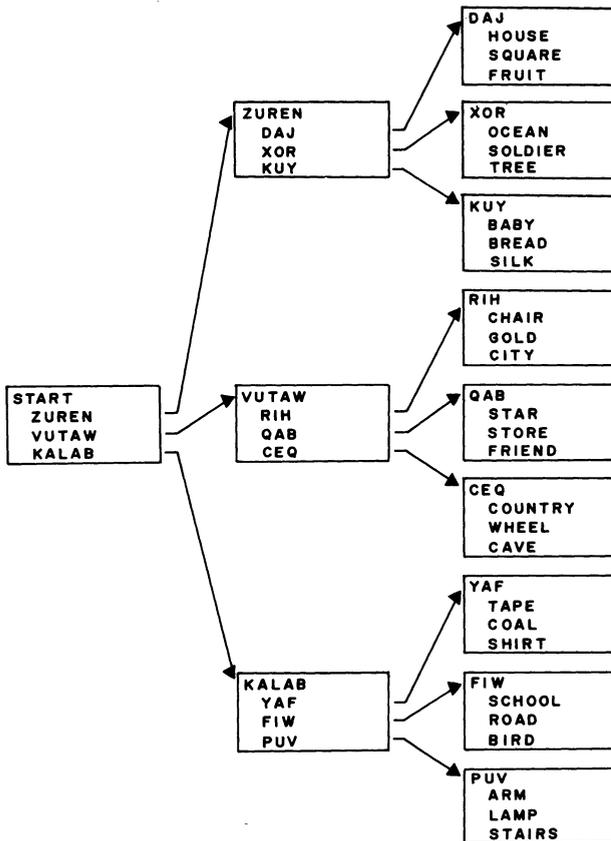


Figure 1. Menu tree showing the 13 frames composed of 3 paralog at Level 1, 9 trigrams at Level 2, and 27 target nouns at Level 3.

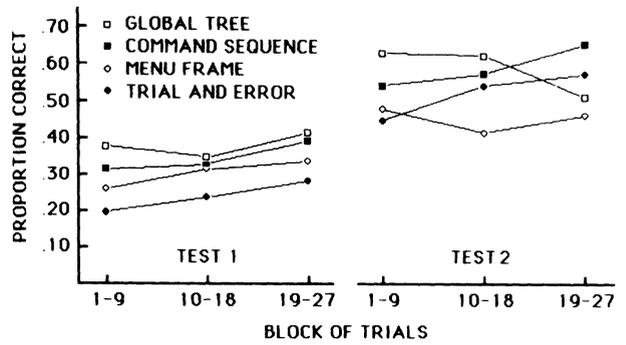


Figure 2. Proportion of targets found across three blocks of nine trials on the Test 1 and Test 2 for each group.

The study-test phases were repeated, and then the subjects were asked to recall and write down all of the paralog, trigrams, and common nouns used in the menu. Finally, subjects were given a menu tree diagram without terms and asked to fill in all menu frames.

RESULTS

Targets Found

Figure 2 shows the proportion of target words located by each group across three blocks of nine trials on Tests 1 and 2. On Test 1 the global-tree group found the highest mean number of targets, followed by the command-sequence group and the menu-frame group. The trial-and-error group found the least number of targets. This effect, however, was not statistically significant [$F(3,56) = 1.32, p > .05$]. Number of targets found increased substantially from Test 1 to Test 2 for all groups [$F(1,59) = 55.43, p < .001$]. On Test 2 the command-sequence and global-tree groups performed equally well. The trial-and-error group was third, and the menu-frame group found the lowest mean number of targets. Again these differences were not significant [$F(3,56) = 1.22, p > .05$]. Finally, it can be seen that the greatest increase from Test 1 to Test 2 occurred for the trial-and-error group; however, the interaction of group x test was not significant [$F(3,56) = .91, p > .05$].

Although the overall number of targets found did not differ significantly with study method, the method of study affected the rate of learning of the menu structure, both within and between tests. Overall, there was an increase in the mean number of targets found across the three blocks (multivariate F block = 3.97, $p < .05$). Generally, performance improved across blocks on both Tests 1 and 2; however, performance by the menu-frame and global-tree groups decreased on Test 2. This effect was indicated by the group x block interaction; however, it did not reach statistical significance (multivariate F group x block = 2.02, $p > .05$).

Location of Errors

Targets might have been missed due to an incorrect choice at the first or second level of the menu. Propor-

Table 1
Proportion of Errors at Level 1 and Errors at Level 2
Given a Correct Choice at Level 1

Group	Test 1		Test 2	
	Level 1	Level 2	Level 1	Level 2
Command Sequence	.47	.39	.70	.20
Menu Frame	.49	.39	.51	.30
Global Tree	.46	.35	.54	.19
Trial and Error	.46	.48	.52	.25
Overall	.47	.40	.52	.23

tion of errors at Level 1 and errors at Level 2, given a correct choice at Level 1, are shown in Table 1.

In all groups subjects made more errors at Level 1 than at Level 2 [$F(1,56) = 108.01, p < .001$]. A significant overall decrease from Test 1 to Test 2 was found [$F(1,56) = 53.23, p < .001$]; however, the decrease in errors was not uniform across level. In fact a slight increase in the proportion of errors occurred at Level 1 from Test 1 to Test 2, whereas there was a substantial drop in the error rate at Level 2. This interaction of test \times level was significant [$F(1,56) = 10.02, p < .01$].

Recall of Menu Terms

Table 2 shows the mean number of terms recalled by each group. For the paralog and trigrams, only the first letter of the term had to be recalled in order to be scored as correct. The command-sequence and the menu-frame groups recalled the greatest number of paralogues used in the main menu, and the trial-and-error and the global-tree groups recalled the fewest [$F(3,56) = 2.77, p < .05$]. At the second level the same ordering occurred; the command-sequence and menu-frame groups recalled the most trigrams, and the trial-and-error group recalled the least [$F(3,56) = 2.98, p < .05$]. The same ordering occurred for recall of the target words; however, this difference was not statistically significant [$F(3,56) = 1.53, p > .05$].

It was expected that performance on the menu system, in terms of the number of targets found, should be correlated with the recall of menu terms. Correlations were computed for each group between the number of targets found on Test 2 and the total number of menu terms recalled. Significant relationships obtained for the menu-frame group ($r = .68, p < .01$) and for the command-sequence group ($r = .91, p < .001$). However, the correlations for the trial-and-error group ($r = .48, p > .05$) and for the global-tree group ($r = .48, p > .05$) were not significant. Furthermore, a comparison between the correlations of the first two groups and the last two was significant ($z = 2.12, p < .05$). This suggests that the type of training on the menu resulted in different acquired representations of the menu tree. For the command-sequence and menu-frame groups, the representation was dependent on recall memory for terms, whereas for the global-tree and trial-and-error groups, it was not.

Reconstruction of Menu Trees

Subjects were asked to reconstruct the menu tree by filling the terms into a diagram such as that shown in Figure 1. Items were counted correct if they were placed as alternatives under the correct term for each frame, irrespective of order. Table 3 shows the mean number of correct terms at Levels 2 and 3 of the tree.

Groups did not differ significantly in placing the correct number of paralogues at Level 2 [$F(3,56) = 1.14, p > .05$] or the correct number of nouns at Level 3 [$F(3,56) = 1.97, p > .05$]. However, the ability to reconstruct the menu tree was significantly related to performance on Test 2. Correlations were particularly high for the trial-and-error group ($r = .81, p < .001$), the global-tree group ($r = .78, p < .001$), and the menu-frame group ($r = .77, p < .001$). For these groups reconstruction performance accounted for 60%-65% of the variance in the number of targets found. The correlation was somewhat less for the command-sequence group ($r = .68, p < .01$), accounting for about 45% of the variance. A comparison between the correlations in the first three groups and the command-sequence group was significant ($z = 2.02, p < .05$). This suggests that although performance did not vary greatly among the four groups in terms of the number of targets found, the groups differed in some fundamental ways of using acquired information about the tree. Subjects in the command-sequence group relied on the rote memory of commands without forming a global representation of the menu tree.

Table 2
Mean Number of Terms Recalled by Each Group

Group	Paralogs* (Max = 3)		Trigrams* (Max = 9)		Nouns (Max = 27)	
	Mean	SD	Mean	SD	Mean	SD
Command Sequence	2.8	.6	6.9	1.5	21.5	4.3
Menu Frame	2.6	.5	5.8	2.3	18.8	3.6
Global Tree	1.9	1.4	4.1	2.4	17.8	7.4
Trial and Error	1.8	1.0	3.8	2.0	13.8	6.4
Overall	2.3		5.2		18.0	

Note—SD = standard deviation. *Only the first letter of the term had to be recalled correctly.

Table 3
Mean Number of Terms Correctly Placed in
the Menu Tree by Each Group

Group	Level 2 (Max = 9)		Level 3 (Max = 27)	
	Mean	SD	Mean	SD
Command Sequence	5.9	2.2	14.0	9.2
Menu Frame	3.8	2.3	13.2	10.0
Global Tree	4.2	2.4	16.1	7.9
Trial and Error	3.8	2.6	15.0	8.9
Overall	4.4		14.6	

Note—SD = standard deviation.

DISCUSSION

The present study used a content-free menu with low-meaningful terms leading to the target nodes in the menu tree. Consequently, extensive training was required for subjects to perform above chance level. Although the four types of training were substantially different, overall search performance differed only marginally among the groups. It was expected that training on the global tree would lead to high stable performance. Although the global-tree group did perform best on Test 1 and on the first two blocks of Test 2, rather than improving across trials as did other groups, the group's performance decreased on Test 2. It is possible that the search for targets presented in a random order interfered with memory of the menu tree.

Trial-and-error training is often encouraged in menu selection systems. Immediate hands-on experience helps to familiarize the user with the system and facilitate performance. In the present case, since a computer system was not used, no facilitation was expected and, indeed, subjects did not perform well initially. However, performance increased substantially over trials, indicating that subjects acquired new information about the menu tree as they were tested.

Training on the sequence of commands is essentially rote learning of associations. It was not expected that subjects would do well studying a list of 27 command sequences; however, subjects in the command-sequence group performed nearly as well as those in the global-tree group. Furthermore, performance increased with trials, indicating that subjects acquired information about the menu tree as they were tested.

Finally, training on the 13 menu frames themselves was not expected to be as beneficial as training on the global tree. This trend was obtained, but the difference was slight.

It is generally assumed that the major advantage of menu selection systems over command languages is that users do not need to recall the exact terms used in the system; they need only recognize correct alternatives. Analyses of recall and reconstruction data suggest that the method of training may have led subjects to learn the menu system in fundamentally different ways. For subjects in the command-sequence and the menu-frame groups, menu performance was contingent upon recall ability. For the trial-and-error and global-tree groups, performance was not significantly related to recall. Moreover, menu performance in the command-sequence group seemed to be less reliant on reconstruction ability, whereas other groups showed significantly higher correlations between menu performance and reconstruction ability. This suggests that subjects in the command-sequence group relied primarily on recall of associations, whereas the other subjects used the tree structure.

A major concern in the use of menu selection systems is the frequency and location of errors. In the present study a much higher error rate was found at Level 1 than at Level 2. There are several factors that may have led to this. Although system designers may build menu trees in a top-down stepwise refinement manner, human associative memory probably proceeds from the target word back up the tree. Once the bottom-up path is generated, the user can proceed down. Because of the remoteness of Level 1 from the target, more errors would be expected at the top level of the tree. One would expect that this difference in errors would change substantially for more meaningful menus in which forward chaining of associations is built into the semantics of the tree.

REFERENCES

- ALLEN, R. B. (1980). *Usage of menus and trees* (Technical Memorandum, No. 80-1359-3). Murray Hill, NJ: Bell Laboratories.
- ARCHER, E. J. (1960, October). A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 74(Whole No. 497).
- BILLINGSLEY, P. A. (1982). Navigation through hierarchical menu structure: Does it help to have a map? *Proceedings of the 26th Annual Meeting of the Human Factors Society*, pp. 103-107.
- MCEWEN, S. A. (1981, May). An investigation of user search performance on a Telidon information retrieval system. *Telidon Behavioural Research 2: The Design of Videotex Tree Indexes*. Ottawa, Canada: Government of Canada.
- MILLER, D. P. (1981, October). The depth/breadth tradeoff in hierarchical computer menus. *Proceedings of the 25th Annual Meeting of the Human Factors Society*, pp. 296-300.
- SHNEIDERMAN, B. (1980). *Software psychology*. Cambridge, MA: Winthrop.
- TAYLOR, J. D., & KIMBLE, G. A. (1967). The association value of 320 selected words and paralogues. *Journal of Verbal Learning & Verbal Behavior*, 6, 744-752.
- THORNDIKE, E. L., & LORGE, I. (1944). *The teacher's word book of 30,000 words*. New York: Teacher's College Press, Columbia University.
- UNDERWOOD, B. J., & SCHULZ, R. W. (1960). *Meaningfulness and verbal learning*. Philadelphia: Lippincott.

(Manuscript received for publication December 11, 1986.)