Predicting the Skilled Use of Hierarchical Menus With the Keystroke-Level Model

David M. Lane, H. Albert Napier, Richard R. Batsell, and John L. Naman
Rice University

ABSTRACT

This article addresses a key question in the application of Card, Moran, and Newell's (1983) keystroke-level model to software in which users specify a command by working through a system of hierarchical menus. For example, to insert a row in Lotus 1-2-3®, the user makes three menu choices: W for worksheet, I for insert, and R for row. In the keystroke-level model, it is assumed that a time-consuming mental operation precedes each command. The question in the application of the keystroke-level model to hierarchical menu systems is whether the keystrokes WIR in the previous example constitute the execution of three commands and thus require three mental operations or whether WIR acts as a single command and requires only one mental operation. Data were collected from four highly experienced Lotus 1-2-3 users as they went about their day-to-day work. Strong evidence that only one mental operation is involved in choosing from a hierarchical menu system was obtained. We hypothesize that the discrepancy of our results from the data of others is due to the fact that our subjects were more experienced. The implications of our findings for the design of menus is discussed.

Authors' present addresses: David M. Lane, Rice University, Houston, TX 77251-1892; H. Albert Napier and Richard R. Batsell, Jones Graduate School of Administration, Rice University, Houston, TX 77251-1892. John L. Naman, Katz Graduate School of Business, University of Pittsburgh, Pittsburgh, PA 15260. E-mail for David M. Lane: DAVIDL62@AOL.COM.
1. INTRODUCTION

Our research focuses on commands issued by users through a series of menu choices. For example, to insert a row using Lotus 1-2-3®, one goes through a series of menus choosing first "Worksheet," then "Insert," and finally "Row." To initiate a command, a user types a backslash. Menu choices are then made in one of two ways: (a) by using the cursor-movement keys to position the cursor over the desired menu choice and then pressing the Enter key and (b) by typing the first letter of each choice (e.g., WIR). Experienced users invariably use the latter method for frequently used commands (Napier, Batsell, Lane, & Guadagno, 1992).

This method of command choice is very common and is by no means limited to Lotus 1-2-3. For instance, in Microsoft® Windows™, a user can employ the mouse to navigate through a series of menus or can type a letter (not necessarily the first) associated with each menu choice. Given the apparent trend in the software industry toward graphical user interfaces, many software companies are revising their products so that command choices can be made by using a mouse or by typing a letter associated with each of a series of menu choices.

In this article, we apply the keystroke-level model (Card, Moran, & Newell, 1980, 1983) to the issuing of commands that require a series of menu choices. The keystroke-level model has two parameters that apply in this situation: $K$ and $M$. $K$ is the time to press a single key on the keyboard and is about 0.28 sec for the average typist; $M$ is the time to mentally prepare for a command and is about 1.35 sec. It is important to distinguish between the time to mentally prepare for a command and planning time, the time to decide that the operation referenced by the command needs to be performed. For instance, the time to mentally prepare for the command WIR does not mean the time to decide that a row should be inserted. Instead, it is the time to prepare to issue the command once it has been decided that a row should be inserted. This preparation involves mainly recalling the name of the command needed to accomplish the task.
According to the keystroke-level model, the first keystroke in a command name contains an M, whereas subsequent letters in the command name do not. For example, for the function command name SUM, there is an M associated with the S but not with the U or the M. The problem in applying the keystroke-level model to commands such as WIR is that it is not clear whether to consider WIR to be three commands and involve three Ms or one command and involve only one M. If WIR is assumed to involve three Ms, then the predicted time to execute it would be $3M + 3K = 4.89$ sec. If it is assumed to involve only one M, then the predicted time would be $M + 3K = 2.19$ sec. Because the former time is more than twice the latter time, the decision about whether there should be one or three Ms is critical.

One of Card et al.'s (1983) rules for determining whether or not to place an M before a keystroke is that there should be no M if the keystroke is fully anticipated. In this context, however, it is difficult to determine whether the I and R in WIR should be considered fully anticipated. If a user retrieves the chunk WIR when wishing to insert a row, then the I and R are fully anticipated just as the U and M in the sum function are. If, however, WIR were treated by the user as three commands, then the second and third commands would not be fully anticipated. Card et al. (1983) acknowledged that it is difficult to determine exactly what constitutes a cognitive unit. For example, they stated: "It is clear that these heuristics do not capture the notion of method chunks precisely, but are only approximations. Further, whether something is 'fully-anticipated' or is a 'cognitive unit' is sometimes ambiguous" (p. 268).

Existing research most relevant to the placement of Ms in commands consisting of multiple menu choices was conducted by J. R. Olson and Nilsen (1988). As part of a larger investigation, J. R. Olson and Nilsen studied the performance of seven full-time students in the University of Michigan MBA program. These students had all used Lotus 1-2-3 either in classwork or as part of their employment before returning to school or both.

J. R. Olson and Nilsen's subjects were tested on a Lotus 1-2-3 command that involves a series of menu choices: Worksheet Column Set-Width (WCS). This command is used to set the width of a column and is followed by digits indicating how many characters wide the column should be. The command is followed by "enter" to indicate that it is finished. Thus, to set the width of a column to 30 characters, one would type "/WCS30<", where "<" is used to indicate "enter."

J. R. Olson and Nilsen applied the keystroke-level model to this sequence of key presses as follows: MK[w]MK[c]MK[s]MK[3]MK[0]MK[Enter]. Thus, J. R. Olson and Nilsen placed an M before each menu choice. Their estimates of M and K were 1.08 sec (95% confidence interval: 0.8 to 1.4) and .20 sec (95% confidence interval: 0.1 to 0.3), respectively. These values are close to the values of 1.35 and 0.28 calculated by Card et al. (1983), thus
supporting the decision to place an M before each menu choice. If an M had been placed only before the first menu choice (W) and before the terminating Enter key, then the estimate of M would have been 2.67, a value almost double Card et al.'s estimate of 1.35. Further supporting the decision to place an M before all menu choices rather than just before the W was the finding that the time to press the W was not longer than the time to press the C or the S.

A detailed look at J. R. Olson and Nilsen's data, however, reveals some uncertainties about whether an M should be placed before each menu choice. Figure 1 shows the time for each key press of the command WCS30<. An inspection of Figure 1 shows that the two slowest keystrokes were the first digit of the column width (3) and the Enter key. It is logical to assume that the additional time taken to enter a 3 represents, in part, the time it took to decide how wide to make the column. As such, it involves some planning time. The time needed to press the Enter key at the end of the sequence is likely to include the time to verify that the command is correct.

Recalculating M using only the keystrokes for which there is no reason to suspect that the processes of planning and verifying are inflating the time (W, C, and S) reduces the value of M from 1.08 to 0.81 sec. It is not clear, therefore, whether a ‘‘full’’ M should be placed before each of the menu choices. The research described here was designed to address directly the problem of placing Ms in commands consisting of a series of menu choices.

Our method was to record the keystrokes and interkey latencies of users of
Lotus 1-2-3. We sought to determine whether the data could be better fit by a model that assumes that an M should precede each menu choice or by a model that does not. To maximize applicability to the real work environment, we analyzed users as they performed their everyday tasks at work using Lotus 1-2-3.

2. EXPERIMENTAL METHOD

2.1. Subjects

Four very experienced users of Lotus 1-2-3 who worked for a Fortune 500 industrial company were the subjects in this study. All subjects used Lotus 1-2-3 at least 2 hr each day in their work activities. According to Card et al.'s (1983) classification scheme, these subjects would be considered "dedicated." The subjects had used Lotus 1-2-3 from 1 to 3.5 years ($M = 2.50, SD = 1.06$), and their average daily use ranged from 2.0 to 2.5 hr per day ($M = 2.13, SD = 0.22$). Clearly, these are expert users.

2.2. Procedure

A customized version of a commercially available keystroke recorder software package called Total Recall™ was installed on each subject's computer at work. The package recorded the subjects' keystrokes and elapsed time between keystrokes while they used Lotus 1-2-3. Each subject used a COMPAQ® 12 MHz 286 personal computer and Lotus 1-2-3 Release 2.01.

The data were collected as the subjects completed their regular work activities over a period of several months. The subjects were professional employees who worked in an accounting department or as an administrative assistant for a department executive.

Two commands that consisted of a series of menu choices were extracted from the keystroke protocols and analyzed. The first, WCS, was chosen so that our data could be compared directly to J. R. Olson and Nilsen's. The second, WIR, was chosen because it is a frequently used Lotus 1-2-3 command that is three keystrokes long.

3. EXPERIMENTAL RESULTS

The first analysis was conducted using commands of the form WCSxx, where $xx$ refers to two digits typed to indicate the desired column width. Thus, commands that set the column width anywhere in the range of 10 to 99 characters were included. Only two of the four subjects issued commands of this form frequently enough to analyze. One subject issued the WCSxx form
of the WCS command 19 times, and the other subject issued the command 28
times; the other two subjects issued the command in this form only one and
zero times. If a subject issued a WCS command two or more times in
succession, only the first of the sequence was analyzed, because J. R. Olson
and Nilsen (1988) had found that response times decrease with successive
executions of a command.

Figure 1 shows the mean times for each keystroke in this sequence for the
subjects. An inspection of this figure shows clearly that the times for each of
the three keystrokes in WCS were so short that they could not possibly include
an M. The mean times for the W, t(1) = 31.86, p = .02; C, t(1) = 24.56,
p = .026; and S, t(1) = 32.71, p = .019, were each significantly lower than
1.28 sec. This means that not only is there no M in the keystrokes following
the keystroke for the first letters in the command, but there is no M in any of
the keystrokes of the command. Evidently, the time prior to pressing the
Backslash key includes the M for the command as well as planning time. The
command WCS is therefore a single cognitive unit, and all the keystrokes in
the command are fully anticipated once the Backslash key is pressed.

To further investigate the dynamics of the WCS command, we examined
all instances of the WCS command whether or not they were of the form
WCSSxx. Included were WCS commands in which subjects used the cursor-
movement keys to specify the column width, in which subjects selected one or
more of the successive menu options by pressing the Enter key rather than
typing the first letter of the menu selection, and in which subjects issued
commands of the form WCSSx.

Three of our four subjects issued WCS commands in sufficient quantity to
analyze (111 times, 82 times, and 91 times; the other subject issued the
command only 4 times). If a subject issued a WCS command two or more
times in succession, only the first of the sequence was analyzed. The mean
times for these three subjects on the W, C, and S commands were 0.15, 0.17,
and 0.20 sec, respectively. These times are all significantly different from
1.28, t(2) = 22.77, p < .01; t(2) = 42.51, p < .01; t(2) = 12.08, p < .01.

When an M is assumed to occur prior to each menu choice, the total time
is M + K or, using J. R. Olson and Nilsen's (1988) estimates of M and K,
1.28 sec. If no M is assumed, then only a K is required and the time is 0.20
sec. The average keystroke times for the subjects in the current study were
extremely close to the times predicted by the version of the keystroke model
that assumes that no M is present prior to each letter. This supports the
assertion that no M is present for any of the letters in the command. The one
M involved apparently occurs before the Backslash key is pressed and is
therefore totally confounded with planning time.

J. R. Olson and Nilsen's times fall between the times predicted based on the
assumption that an M occurs before each menu choice and the times predicted
based on the assumption that no Ms occur.
We also collected data for the command WIR. There were three subjects who issued the command frequently enough to analyze (63, 48, and 38 times; the other subject only issued it 6 times). If subjects issued the command more than once in a row, only the first occurrence was analyzed. The mean times for the three keystrokes were 0.186, 0.310, and 0.164 sec, respectively. All three times are significantly lower than 1.28, $t(2) = 41.66, p < .01; t(2) = 25.70, p < .01; t(2) = 117.00, p < .01$. As in the command WCS, the times for the keystrokes in WIR are far too short to include an $M$. These data, therefore, further support the conclusion that a command is a single cognitive unit that is fully anticipated once the Backslash key is pressed.

To test whether the difference between our data and J. R. Olson and Nilsen's is due to differences in skill with particular tasks or a more general difference, we looked at the performance of the two users whose data are included in Figure 1 on less frequent spreadsheet commands. A random sample of times on three-letter commands previously found to be infrequently used (Napier et al., 1992) was taken, eliminating repeated commands. Between the two users, 44 commands were considered. The mean times for the three keystrokes of these commands were 0.35, 0.38, and 0.57 sec. Although these times are shorter than the 1.28 sec expected for an $M + K$, they are about twice as long as the times of 0.18, 0.19, and 0.26 sec shown on the frequently used WCS command.

4. DISCUSSION

The present data indicate clearly that, for dedicated users, commands issued through a series of menu choices involve only one $M$ rather than an $M$ for each menu choice. The discrepancy between our results and those of J. R. Olson and Nilsen, who concluded that an $M$ is involved with each menu choice, is most likely due to differences in the expertise of the subjects. Our subjects used Lotus 1-2-3 on a daily basis for years, whereas the subjects in the J. R. Olson and Nilsen study were experienced but not dedicated users.

Our data clarify the need for considering the experience of the user in deciding whether or not to include a mental before each menu choice. For dedicated users, it appears that no mental is involved. Experienced but not yet dedicated users are apparently in a transition state between requiring a mental at each choice and only requiring one mental. Therefore, we suggest a modification to J. R. Olson and G. M. Olson's (1990) conclusion that the time to retrieve a command part is about 1.1 sec. We believe that it would be more accurate to say that the time to retrieve a command part ranges from 0.00 sec to 1.35 sec, depending on the experience of the user with the specific command. Card et al. (1983, p. 282) showed that $M$s vary in time as skill develops. Thus, as users develop skill with a specific command, $M$s decrease
and eventually approach zero for the keystrokes associated with the individual menu choices for that command.

The issue of whether or not to include an $M$ for each menu choice has important implications for interface design. For example, it has implications for whether menus should be wide and shallow (many choices at each level of the hierarchy and therefore few levels) or narrow and deep (few choices at each level of the hierarchy and therefore many levels). If each menu choice requires an $M$, then wide shallow menus would be a necessity because $M$s are of relatively long duration and because shallow menus require fewer $M$s. On the other hand, if there is only an $M$ at the beginning of the sequence, each additional depth of a menu for a well-learned command would only add a $K$. Because $K$s are of relatively short duration, increasing the depth in order to satisfy other design needs such as display clutter would have only a small cost. The depth versus breadth tradeoff is very complex and only partially understood (Norman, 1990), and our results should not be taken to mean that depth is to be preferred to breadth. However, they should inform the software designer about what the tradeoff can be expected to be for dedicated users. In particular, the designer can be confident that a dedicated user can go through a deep menu hierarchy very quickly.

REFERENCES


**HCI Editorial Record.** First manuscript received December 10, 1990. Revisions received September 16, 1991, and April 20, 1992. Accepted by Thomas Moran. Final manuscript received October 1, 1992. — Editor