Behavior Research Methods, Instruments, & Computers
1986, 18 (2), 129-134

SESSION III
SOFTWARE APPLICATIONS

Paper Session
Stephen Link, *Presider*

Remembering computer command names: Effects of subject generation versus experimenter imposition

MARC M. SEBRECHTS, CHARLOTTE T. FURSTENBERG, and ROXANNE M. SHELTON
Wesleyan University, Middletown, Connecticut

Two experiments analyzed the effectiveness of goal statements in aiding recall of self-generated as opposed to experimenter-imposed command names. Subjects were presented with a series of before-after pairs representing the computer states before and after a command was executed. In Experiment 1, during study, one group of subjects generated a command name in response to each pair; a second group generated a goal statement describing the goal to be accomplished in addition to generating a command name. During recall, half of each group was required to recall the name, whereas the other half was required to describe the goal before attempting to recall the name. In Experiment 2, during study, command names (and goals for those subjects in the goal condition) were imposed by the experimenter rather than generated by the subject. Subjects who generated goals and names recalled more command names than did those who generated only names or who received imposed goals and/or names. Generation of an appropriate goal at study improved encoding by helping subjects to select more appropriate command names; generation of an appropriate goal at test improved retrieval for appropriate names only, presumably by activating a relevant subset of names. Even in the relatively simple task of naming and remembering command names, having an appropriate model of the domain through the use of specific goal statements substantially improved performance.

Using computers requires a precise way to specify the tasks that the computer is to perform: this constitutes what is called the command language problem. There are, of course, many alternative ways to implement a command language, including specific command terms, a natural language interface, icons, and menu selection.

This paper focuses on the use of specific command names, a common practice in most current computer systems. The main concerns here are: (1) does self-generation provide a usable set of command names, and (2) what characteristics are important for optimal command generation and recall.

Previous studies have suggested a number of problems with having novice subjects generate command names (see Black & Sebrechts, 1981). In a study by Landauer, Galotti, and Hartwell (1983), subjects were asked to provide a procedure for a series of editing changes. There was no substantial agreement across subjects on basic names for operations. Furthermore, many subjects did not even use the same verb for two different instances of the same type of correction.

In a second study, Landauer et al. (1983) compared the learning of a simplified text editor using one of three types of command names: the modal ("natural") responses from the previous generation study, standard UNIX ED commands, and random names. They found no reliable differences in time to complete a basic set of tasks based on the type of command name, suggesting that the natural naming does not produce more memorable names.

In a related study, Black and Moran (1982) had a group of subjects generate single-sentence descriptions of the difference between two before-after text fragments, representing eight editing operations. They then compared performance of another group who learned the frequently produced names from the generation study with a group that learned designer-generated names. They found no retention differences beween the subject-generated and designer-generated names.

This research was supported by NSF Grant IST82-17572 and by a research grant from Digital Equipment Corporation. The views expressed are those of the author and are not necessarily endorsed by Digital Equipment Corporation. Send correspondence to: M. M. Sebrechts, Department of Psychology, Wesleyan University, Middletown, CT 06457.

These results again suggest that natural command generation by a group of relatively computer-naive subjects is not an effective strategy for designing a command language. They do not, however, indicate whether or not an individual would be good at selecting his or her own command names. There are several studies that suggest that self-generation of names would in fact lead to greater memorability.

Slamecka and Graf (1978) found a retention advantage for self-generated words over the same words that were merely read. This was true for a relatively wide range of presentation conditions, using recall, recognition, and confidence-rating measures. This effect appears to be due to more than simple elaboration, since the same result was later obtained even when the group that received imposed names was given a task that required describing whether a word fit the specified context (McFarland, Frey, & Rhodes, 1980).

The present study attempts to extend these generation results to command naming. The data on command generation do not directly address the possibility of self-generation. And the results on self-generation do not directly address the question about memory in a complex, unfamiliar, but meaningful domain such as command generation.

One of the potential difficulties in generating commands, however, is that subjects do not necessarily have a clear representation of the task. In fact, one of the problems that Black and Moran (1982) observed is that subjects generated general, vague command names. So the effectiveness of generating names may depend upon having a clear specification of the function that is being named. Based on work in other domains (Ausubel, 1960; Mayer, 1975), it is reasonable to conjecture that a goal structure may enhance either the generation or retention of command names. In other words, the effectiveness of self-generation may depend upon the learning context.

In order to test whether encoding or retrieval would be enhanced by specifying a context, the effect of a goal statement at study and at test was also analyzed.

EXPERIMENT 1

Experiment 1 was designed to investigate how well self-generated command names would be recalled and how recall would be influenced by the use of goal statements during command generation and during recall.

Method

Subjects. Fifty-three college students, who had little or no previous computer or word processing experience, were paid $3 for participating in this experiment. Five subjects' data were eliminated from the final analyses because they had not followed directions. The 48 remaining subjects were grouped equally into the four conditions described below.

Materials. The study-session booklets contained 24 before-after pairs in random order. Each pair consisted of a picture of two adjacent computer screen displays: the first indicated how the screen appeared before a command was typed; the second indicated how the screen appeared after the command was executed. When relevant, the contents of the computer memory was displayed in a separate box, marked MEMORY, below each screen. The change between the before and after display was underlined in red to draw attention to the difference. The following is an example of a before-after pair for the INSERT command.

BEFORE	AFTER
A watched pt never boils.	A watched pot never boils.

The test booklets contained the same 24 before-after computer stimulus pairs used in the study session in random order.

Procedure. Each subject was given a pretest questionnaire to assess level of computer and text-editing experience. Subjects were then randomly assigned to one of two study conditions. Subjects in the name-study condition were instructed to generate a command name for each stimulus pair, using no name more than once. Subjects in the goal-plus-name-study condition were instructed to generate a goal statement for each computer pair, explaining what function a computer user would be trying to accomplish, and to generate a command name without using any name more than once.

After completing the study task, subjects were given a distractor test on rating the similarity of dot patterns. Following the distractor task, subjects were given the test booklets with the 24 before-after pairs received during the study condition. Half of the subjects in each study condition were assigned to one of two test conditions. Subjects in the name-test condition were instructed to recall the name that they generated for each stimulus pair. Subjects in the goal-plus-name-test condition were instructed to recall the goal and name (for those in the goal-plus-name-study condition) or to generate a goal and recall the name (for those in the name-study condition).

When subjects had completed the test section, they were paid and debriefed.

Results

Ratings. Preliminary analyses were conducted to characterize the goals and names that subjects generated. Names were rated for frequency using the Kučera and Francis (1967) norms. For each stimulus pair, generated names were divided into natural groupings with lower and higher frequency. When there were no natural groupings, names with a frequency lower than the median of all names generated to that pair were called "low-frequency" names, and names with a frequency higher than the median group were called "high-frequency" names.

In order to determine name appropriateness, a separate rating study was conducted. The command names generated were shown with the related before-after pair, and a new group of subjects was asked to provide an appropri-

ateness rating on a scale from 1 to 9. Names with a rating lower than the median for each stimulus pair were called "low-appropriateness" names; names with a rating higher than the median were called "high-appropriateness" names.

The goals were rated independently by three judges and were classified as good, vague, or incorrect. Good goals appropriately described the action. Vague goals, although correct, were not sufficiently specific to differentiate a specific command. Incorrect goals did not accurately describe the command functions. Since there were few incorrect goals, vague and incorrect goals were combined to form a category called "poor goals."

Recall. Subjects generated names which varied substantially in frequency and appropriateness. However, the largest category of generated names was low-frequency, high-appropriateness names (45%), and the percentage of names in that category was roughly equal across conditions. The mean command recall by name frequency and appropriateness category is shown in Table 1.

Subjects recalled a higher percentage of high-appropriateness names than low-appropriateness names [$F(1,44) = 14.81$, $p < .001$]. There was no effect of name frequency on recall and no interaction between name frequency and name appropriateness.

The mean number of command names correctly recalled in each study and test condition is shown in Table 1. A 2×2 analysis of variance indicated marginal significance of the study condition [$F(1,44) = 3.94$, $p = .054$]. Subjects who generated goals and names at study recalled more names than did subjects who generated only names. There was no interaction effect between study and test conditions.

The reason for the effect of study goal was analyzed more closely by comparing performance after generation of good study goals (i.e., those that describe the specific task) with performance after generation of poor study goals (i.e., either vague or incorrect). As shown in Table 2, generating good study goals resulted in reliably better recall than did generating poor study goals [$F(1,22) = 4.93$, $p < .05$]. The advantage for the goal-plus-name-study condition over the name-study condition is mainly due to the good study goals. Names generated with good study goals resulted in better recall (76%) than did names generated in the absence of goals, whereas names generated with poor study goals (64%) and names generated without goals (62%) were recalled at roughly the same level.

Table 1
Mean Percent Name Recall by Study, Test, Name Appropriateness, and Name Frequency in Experiment 1 (Name Generation)

Condition:	Name Frequency and Appropriateness				
Study/Test	LL	LH	HL	HH	Mean
N/N	58	76	57	75	68
N/G+N	55	63	63	62	57
G+N/N	52	80	61	87	75
G+N/G+N	68	78	50	69	72
Mean	58	74	58	73	

Note—First letter under Condition refers to study, second to test: N = name; G+N = goal-plus-name. L = low; H = high: first letter refers to frequency, the second to appropriateness. Row and column means are weighted by frequency of occurrence in each cell.

Table 2
Mean Percent Name Recall by Test and Study Goals in Experiment 1 (Name Generation)

Condition: Study/Test	Good	Poor
	Study Goal	
G+N/N	79	69
G+N/G+N	74	63
	Test Goal	
N/G+N	60	47
G+N/G+N	76	61

The type of test goal also influenced name recall, as shown in Table 2. Significantly more names were recalled when good test goals were produced than when poor test goals were produced [$F(1,22) = 7.47$, $p < .05$].

Discussion

The main characteristic of command names that affected recall in this study was appropriateness. Once that variable was taken into account, there was no residual effect of relative name frequncy. Although previous research has found that subjects tend to generate high-frequency general names (Black & Moran, 1982), in this experiment, low-frequency high-appropriateness (i.e., task-specific) names were most frequent. This may be the result of the fact that subjects had a wider range and a greater number of functions for which to find names.

Generating goals before generating a command name also increased recall. One possible explanation is that goals simply result in more appropriate names. However, in general, that is not the case. An alternative explanation is that goal statements increase the amount of elaboration (Anderson & Reder, 1979) or the degree of depth of encoding (Craik & Lockhart, 1972; Craik & Tulving, 1975). However, according to this account, goals in general should improve recall, and again there is no uniform effect of goal generation on recall. Instead, goals tended to improve performance only if they were good (i.e., appropriate and specific to the task).

A simple quantitative account of the effect of goals seems inappropriate. Instead, the data indicate that the quality of the study goal is relevant to recall. Generating a goal, when it is correct and specific, helps to select an appropriate name. When a good goal was generated, 80% of the associated names were highly appropriate. When a poor name was generated, only 50% of the associated goals were appropriate.

The quality of the test goal also affected recall. When subjects produced a good test goal, they recalled 68% of the names; when they produced a poor test goal, they recalled only 54% of the names. The presence of a good

goal may help decrease the search space in memory and thus increase the probability of retrieving the correct name.

EXPERIMENT 2

Experiment 1 examined subjects' abilities to remember command names that they generated. Experiment 2 was designed to provide a way to compare recall of an imposed set of names to recall when subjects generate their own names. By controlling the frequency and appropriateness of goals and names, it should be possible to provide a more precise characterization of the properties that influence recall.

Method

Subject. Ninety-six college students, with little or no prior computer or text-editing experience, participated in this experiment. None of them had participated in the name-generation or rating experiments. The subjects were assigned equally to one of four study and test combinations: name-study, name-test; name-study, goal-plus-name-test; goal-plus-name-study, name-test; or goal-plus-name-study, goal-plus-name-test.

Materials. The study-session booklets contained the 24 randomly ordered before-after pairs used in Experiment 1. In the name-study condition, a command name was presented with each stimulus pair. In the goal-plus-name-study condition, both a command name and a goal phrase were presented with each stimulus pair. The command names and goal statements belonged to one of four categories: low frequency and low appropriateness, low frequency and high appropriateness, high frequency and low appropriateness, and high frequency and high appropriateness. Each booklet contained equal numbers of command names in each of the four categories. Across subjects, each stimulus pair appeared equally often with names from each category. For the goal-plus-name-study condition, each goal statement contained a key word belonging to one of the four goal groups: low frequency, low appropriateness; low frequency, high appropriateness; high frequency, low appropriateness; and high frequency, high appropriateness. Each booklet contained equal numbers of goals in each category. Across subjects, each stimulus pair occurred equally often with goals from each category.

The dot-pattern distractor task from Experiment 1 was used. The test booklet contained the 24 before-after pairs from the study session without names or goals.

Procedure. The experiment was run in groups of 1 to 4 subjects. Each subject was given a pretest questionnaire to assess levels of computer programming and text-editing experience. Subjects were then randomly assigned to one of two study conditions. Subjects in the name-study condition were presented with a command name accompanying each of the 24 before-after stimulus pairs. Subjects rated the command for each pair on a 9-point scale with 1 being *highly inappropriate*, 5 being *neutral*, and 9 being *highly appropriate*. Subjects in the goal-plus-name-study condition were given a goal and a name accompanying each before-after stimulus pair. They rated appropriateness of goals and names on the same 9-point scale. The distractor task and test conditions were the same as those in Experiment 1.

Results

Overall, subjects' recall in all four conditions (mean = 53%) was worse than in Experiment 1 (mean = 68%). As shown in Table 3, both goals at study [$F(1,92) = 13.36$, $p < .001$] and goals at test [$F(1,92) = 3.91$, $p < .05$] decreased recall compared to conditions with only names.

There was a marginally significant interaction effect between study and test conditions, indicating that the name-study, name-test condition resulted in better recall than did the other three conditions [$F(1,92) = 3.11$, $p = .08$].

Table 3 also shows the mean name recall broken down by name frequency and name appropriateness across study and test conditions. High-appropriateness names were more likely to be recalled than low-appropriateness names [$F(1,92) = 84.90$, $p = .001$]. There was no significant effect of name frequency, and there were no statistically significant interactions between name frequency and name appropriateness.

The effect of the frequency and appropriateness of the study goal are shown in Table 4. Highly appropriate study goals tended to produce better recall than did less appropriate study goals [$F(1,46) = 3.65$, $p = .06$]. There was no effect of study goal frequency and no interaction between goal frequency and goal appropriateness.

Since test goals were generated by subjects, they were rated for goodness as in Experiment 1. As shown in Table 5, when good test goals were produced, more names were recalled than when poor test goals were produced [$F(1,46) = 25.35$, $p < .001$].

Table 3
Mean Percent Name Recall by Study, Test, Name Appropriateness, and Name Frequency in Experiment 2 (Name Imposed)

Condition:	Name Frequency and Appropriateness				
Study/Test	LL	LH	HL	HH	Mean
N/N	55	72	59	71	64
N/G+N	41	69	44	56	53
G+N/N	40	58	39	53	48
G+N/G+N	38	53	36	60	47
Mean	44	63	45	60	53

Table 4
Mean Percent Name Recall by Test Condition, Study Goal Appropriateness, and Study Goal Frequency in Experiment 2 (Name Imposed)

Condition:	Goal Frequency and Appropriateness			
Study/Test	LL	LH	HL	HH
G+N/N	41	51	49	49
G+N/G+N	45	48	42	52

Table 5
Mean Percent Name Recall by Study Condition and by Test Goals in Experiment 2 (Name Imposed)

Condition:	Test Goals	
Study/Test	Good	Poor
N/G+N	58	43
G+N/G+N	56	40

Discussion

Consistent with the results of Experiment 1, appropriateness of goals and names produced better overall performance, whereas relative frequency had no reliable effect. Appropriateness is therefore important for recall of experimenter-selected (imposed) names as well as for recall of subject-generated names.

Overall performance in Experiment 2 was worse than that in Experiment 1. This is due to the fact that subjects were given a balance of appropriate and inappropriate goals and names. When subjects were provided with command names, they performed as well as when they generated commands of equal appropriateness. Comparing the results for the name-study, name-test condition across the two experiments (Tables 1 and 3) shows that performance was roughly equivalent for the four frequency and appropriateness combinations.

In contrast with Experiment 1, where goals tended to improve recall, the presence of goals in Experiment 2 hindered performance. This result can be attributed in part to the fact that subjects were given both appropriate and inappropriate goals. However, even when the goals were appropriate, performance did not reach the levels of recall found in Experiment 1. This appears to be due to an experimental dissociation between goals and names which eliminates the encoding and retrieval cue effects.

In Experiment 1, study goals improved recall by helping subjects to generate more appropriate names which were personally salient. In Experiment 2, experimental manipulation eliminated the link between study goals and names since subjects did not generate their own names. Even when study goals and test names were appropriate, recall in Experiment 2 was less than that in Experiment 1.

Goals at test likewise had a different effect on recall. In Experiment 1, good test goals improved performance for both appropriate and inappropriate names. Generating a good test goal presumably serves to delimit the subject's search space, which includes names that are personally relevant, even though other subjects may not consider such names appropriate (as measured by appropriateness ratings). In Experiment 2, good test goals helped only in the recall of appropriate names. Since the names are imposed, inappropriate names are likely to fall outside of the individual's search space specified by the goal.

In brief, goals increase recall by increasing access to the names. The different effects of self-generated as opposed to imposed goals and names can be attributed to the "personal reference" of self-generated terms (McFarland et al., 1980). That personal character in turn probably functions by reducing the arbitrariness of relationships (Stein, Littlefield, Bransford, & Persampieri, 1984).

SUMMARY AND CONCLUSIONS

Contrary to suggestions arising from previous research, people can select appropriate command names for a number of computer functions, and in a number of cases, such generated names can be more memorable than imposed names (at least in the short term). At the same time, careful design can produce names that seem to have memorability at least as great as that of self-generated names.

In the present study, when people generated goals and names for functions, they recalled more names than did people who only generated names or who received imposed names or imposed goals-plus-names. Presumably, the goal generation constrained the naming context. Systems that include the capacity to change command names should likewise constrain the context of naming to enhance memorability, perhaps by forcing subjects to answer some small set of questions whenever a new command name is selected.

At the same time, the presentation of a goal at test can interfere with recall. When command names are learned in the context of one goal and are tested with another goal, the second goal can produce interference by activating the wrong associations. In designing documentation, careful attention should be given to the mapping between explanatory statements and associated functions. Although the current data assess this relationship only for command naming, given the generality of the underlying processes, the results are likely to generalize to other aspects of computer skill acquisition.

REFERENCES

ANDERSON, J. R., & REDER, L. M. (1979). An elaborative processing explanation of depth of processing. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory*. Hillsdale, NJ: Erlbaum.

AUSUBEL, D. P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology*, **50**, 267-272.

BLACK, J. B., & MORAN, T. P. (1982). Learning and remembering command names. In *Proceedings of human factors in computer systems* (pp. 8-11). Washington, DC: Association for Computing Machinery.

BLACK, J. B., & SEBRECHTS, M. M. (1981). Facilitating human-computer communication. *Applied Psycholinguistics*, **2**, 149-177.

CRAIK, F. I. M., & LOCKHART, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, **11**, 671-684.

CRAIK, F. I. M., & TULVING, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, **104**, 268-294.

KUČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

LANDAUER, T. K., GALOTTI, K. M., & HARTWELL, S. (1983). Natural command names and initial learning: A study of text-editing terms. *Communications of the ACM*, **26**, 495-503.

MAYER, R. E. (1975). Different problem-solving competencies established in learning computer programming with and without meaningful model. *Journal of Educational Psychology*, **67**, 725-734.

MCFARLAND, C. E., FREY, T. J., RHODES, D. D. (1980). Retrieval of internally versus externally generated words in episodic memory. *Journal of Verbal Learning & Verbal Behavior*, **19**, 210-225.

SLAMECKA, N. J., & GRAF, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Human Memory*, **4**, 592-604.

STEIN, B. S., LITTLEFIELD, J., BRANSFORD, J. D., & PERSAMPIERI, M. (1984). Elaboration and knowledge acquisition. *Memory & Cognition*, **12**, 522-529.