THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 62, No. 6, July–August 1983
Printed in U.S.A.

Human Factors and Behavioral Science:

On Abbreviating Command Names

By L. A. STREETER,* J. M. ACKROFF,† and G. A. TAYLOR†

(Manuscript received May 21, 1982)

Test subjects' abbreviations of command names and randomly selected English words were examined for production regularities. Abbreviation rules based primarily on a word's number of syllables were devised to capture regularities observed in people's productions. This rule set was compared to two simpler abbreviation rules—vowel deletion and truncation. In subsequent learning experiments, separate groups of subjects learned the rule-derived abbreviations for words, while other groups learned the most frequently given abbreviation for each word. Subjects who studied rule-derived abbreviations remembered substantially more of them when prompted with full words than did subjects who studied the most frequently given abbreviations. Moreover, the rule-based abbreviations were superior even for those for which the rule-produced and the most frequently produced abbreviations were identical. When the task was reversed (recall the source term given an abbreviation), performance was best for vowel deletion abbreviations and worst for the rule set abbreviations. We suggest that both memorability of abbreviations and the probability that people will spontaneously produce a "correct" abbreviation are increased by: (1) selecting abbreviations using a vowel deletion rule for one-syllable words and an acronym rule for multiple-word terms, as well as (2) allowing variable length truncations of words.

* Bell Laboratories. † American Bell.

I. INTRODUCTION*

Zipf[1] observed that in a number of languages frequently used words tend to be short. Further, this relation appears to be a causal one: as a word or phase increases in usage, it becomes shorter. Examples of word or phrase abbreviation are plentiful in English (e.g., "television" becomes "TV"). Abbreviating is particularly prevalent in government, corporations, and disciplines sharing a specialized vocabulary. Often the abbreviation process is so complete that the antecedent term is forgotten and only the abbreviation is retained (e.g., "cathode ray tube" becomes "CRT" or "light amplification by stimulated emission of radiation" becomes "laser"). Given the frequency and importance of this phenomenon, it seems reasonable to ask whether the process is systematic. That is, given a word or phrase, can one predict how it most likely will be abbreviated? If one could formulate abbreviation rules rooted in natural abbreviating behavior, would these abbreviations be "better" than abbreviations not suggested by behavioral data?

Discovering ways to generate abbreviations that are easy to learn and use has practical significance, since in many circumstances brevity is needed or desired. For instance, while infrequent users of a computer command language may be content to enter complete command names and arguments, frequent users often prefer terse command strings. To avoid typing long command strings, mechanisms should exist for abbreviating commands and their arguments.

There are some common ways to provide abbreviation facilities. One, truncation, allows the user to enter only as much of the command name (from the beginning to some point less than the whole word) as necessary to differentiate it from all other commands in the language. Another provides the user with standard system abbreviations. It is also possible to use both of these methods.

Selecting "good" abbreviations is a complicated problem. There is probably no single measure of an abbreviation's goodness, since there are some competing requirements to consider. For instance, one might want abbreviations in a system to be unique, easy to type, memorable, and natural (i.e., users who knew only the command would be more likely to produce this particular abbreviation than any other). However, some of these requirements are at odds with one another. Consider "easy to type" and "memorable." As Experiment II shows (see Section III), abbreviations produced by a simple vowel deletion rule are easy to remember or generate. Thus, given the command "journal,"

---

* Some of the material contained in this paper was presented at the 52nd Annual Meeting of the Midwest Psychological Association, St. Louis, May 1980 and the 52nd Meeting of the Eastern Psychological Association, New York City, April 1981.

subjects can generate the requisite vowel deletion abbreviation, "jrnl," with nearly perfect accuracy. However, this abbreviation is probably not easy to input, since the user must spell the whole word and omit every vowel. While we have no production data, it is likely that using a vowel deletion abbreviation rule for long words is no faster than entering the complete word. In fact, it may be slower in many cases.

Consider "easy to type" and "unique." Truncation produces abbreviations that seem easy to type (at least no more difficult to type than the whole word for a skilled typist and probably easier than the whole word for a nonexpert typist), particularly if one truncates to only a few letters. However, the likelihood of collisions among commands with truncation increases as the number of letters used decreases. If one requires uniqueness of abbreviations and has a truncation rule, the user must learn the minimum number of letters required to distinguish a particular command from all other commands. An additional burden may be imposed on the user if the command lexicon grows to include words whose truncations suddenly conflict with old abbreviations. If the lexicon is large enough, conflicts between abbreviations are bound to arise. One way to resolve conflicts in a rule-based system is to allow exceptions to the abbreviation rule. However, to what degree rule-generated systems can tolerate exceptions is an open issue worthy of future investigation. That is, at what point do exceptions to the rule render the rule worthless? It could well be that rules are useful only when there are relatively few or even no exceptions.

While each of these abbreviation requirements is important and critical from a design standpoint, the present study focuses on abbreviation naturalness and memorability. We examined abbreviations to determine how much regularity existed in those produced naturally. Next we incorporated the observed regularity into a set of abbreviation rules and compared the learning of rule-based abbreviations to the learning of the most frequently produced abbreviations for each word. The learning process was also reversed to include recalling of a word from its abbreviation, when the abbreviation was produced either by rules or by consensus.

II. EXPERIMENT I: REGULARITIES IN NATURALLY PRODUCED ABBREVIATIONS

2.1 Introduction

To determine how people abbreviate, we solicited abbreviations for a number of command terms. We then looked for regularities and the amount of agreement in these productions.

There were already data indicating that long and short words are abbreviated differently. Hodge and Pennington[2] examined how people

abbreviate words that varied in letter length. They observed that subjects used two principal mechanisms. One method, "contraction" (omitting word internal letters), tended to be used to abbreviate short words, whereas the other method, "truncation," was favored in abbreviating long words. However, they did not examine other linguistic properties of words associated with each of these strategies.

To examine abbreviations, we partitioned them into classes based on whether a particular abbreviation was an instance of a vowel deletion rule, a truncation rule, or an acronym formation rule. (Here acronyms are formed from multiple words, in which the first letter of each word is taken for the abbreviation; e.g., "cathode ray tube" becomes "CRT.") We examined whether the proportion of abbreviations accounted for by each rule varied with the number of syllables or words in the abbreviated term.

2.2 Method

2.2.1 Stimuli

The vocabulary consisted of 81 command names and arguments that were to be used in a large computer system. Many of these terms, such as move, copy, and insert, are common to other computer systems. Others were unique to this particular application, such as usage billing number. The items varied from one to three words. Table I gives statistics for the commands and arguments used in the present study in terms of number of words and syllables and average number of vowels and consonants.

2.2.2 Subjects

There were two groups of subjects; one group consisted of 19 psychologists doing human factors work at Bell Laboratories, Holmdel, New Jersey, while the other set consisted of 30 adult females living in the Holmdel area. The Bell Laboratories employees volunteered their time; the other subjects were paid for participating.

2.2.3 Procedure

We gave each group a randomized list of the command terms and asked them to produce "a good abbreviation" for each, that is, one that would be easy to use and remember. Because the command list grew, the human factors psychologists saw a subset of 42 terms, whereas the other subjects saw 81 terms. The paid subjects produced the abbreviations and then performed a 30-minute text-editing task. After the intervening task, these subjects recalled as many of the command terms as they could. There was no free recall task for the psychologists.

2.3 Results and discussion

To determine the consistency with which subjects produced abbreviations, we examined the abbreviation that was most frequently produced for each item. Overall, the concurrence across the 81 terms was 37 percent; that is, on the average the most commonly produced abbreviation for each item comprised 37 percent of the abbreviations. There were no discernible differences between the psychologists and local residents in their abbreviations of the terms. The mean percentage concurrence was identical to the nearest percent for the two groups. The groups did not differ on other production measures, such as number of different abbreviations produced or the most commonly produced abbreviation. Consequently, data for the two groups were pooled in subsequent analyses.

Table II shows the average concurrence across items and the standard deviation both for all terms combined and separately by number of syllables. Concurrence decreased markedly as the number of syllables in a word increased and, correspondingly, the number of different abbreviations given for each term increased substantially. Thus, ab-

Table I—Composition of command names and arguments

	Overall N = 81	One-Syllable Words N = 15	Two-Syllable Words N = 42	Three-Syllable Words N = 12	Four-Syllable Words N = 2	Multiple Words N = 10
Mean number of:						
Consonants	4.11	2.60	3.80	4.58	4.50	7.10
Vowels	2.77	1.67	2.55	3.17	4.50	4.50
Syllables	2.19	1.00	2.00	3.00	4.00	3.50
Words	—	—	—	—	—	2.10

Table II—Experiment 1: Abbreviation production

	Overall N = 81	One-Syllable Words N = 15	Two-Syllable Words N = 42	Three-Syllable Words N = 12	Four-Syllable Words N = 2	Multiple Words N = 10
Mean percentage concurrence	37%	56%	36%	32%	28%	24%
Standard deviation	(16%)					
Mean number of different abbreviations	11.95	5.73	11.13	16.00	18.00	18.90
Standard deviation	(5.68)					
Average number of terms recalled	3.25	4.87	3.20	2.50	2.00	2.20
Standard deviation	(3.15)					

breviation production was more homogeneous in simpler linguistic environments.

The one or two most frequently elicited abbreviations for each term were studied to determine whether abbreviating could be characterized by some "rule-governed" process. If one ignored the linguistic composition of the words, the abbreviation process appeared unorderly. However, if the terms were partitioned into three classes (monosyllabic, polysyllabic, and multiple words), there was some consistency within each class. By inspection there appeared to be three different rules operating in people's productions—vowel deletion (delete word internal vowels), truncation, and "acronym formation" (for multiple-word terms, select the first letter of each word).

We derived rules to account for the regularity observed in the abbreviating behavior. Table III shows the set of rules that appear to describe the data. The application of a particular rule depended on the number of syllables and/or number of words in the command term.

Table IV compares the proportion of abbreviations accounted for by the three different rules (truncation, vowel deletion, and the rule set from Table III) with the most frequently given abbreviation from each term ("popular"). We combined all truncations from the first letter to one less than the number of letters in a word. We defined vowel deletion as the deletion of all vowels following the first consonant in a word ("a," "e," "i," "o," "u," and "y," when it functioned as a vowel). It is important to remember in interpreting Table IV that popular abbreviations and those produced by the three rules are *not necessarily mutually exclusive.* For example, for one-syllable words there is a great deal of overlap between the rule set and vowel deletion. The rule set for one-syllable words is essentially a vowel deletion rule with only a few other features. However, there was no overlap between the truncated one-syllable abbreviations and the rule set or the vowel

Table III—Abbreviation rules

- Monosyllabic words:
  1. Take initial letter of the word and all subsequent consonants.
  2. Make adjacent double letters single.
  3. If more than four letters remain, retain the fifth letter if it is part of a functional cluster (such as *th, ch, sh, ph, ng*); otherwise, truncate from the right.
  4. Delete the fourth letter if it is silent in the word.
- Polysyllabic words:
  1. Take the entire first syllable.
  2. If second syllable starts with a consonant cluster, add it.
  3. If first syllable is a prefix (such as *de, re, in*) add the second syllable.
  4. Make final double consonants single.
  5. Truncate to four letters (but always retain entire first syllable).
- Multiple words:
  1. Retain the first letter of each word as the abbreviation.

Table IV—Proportion of abbreviations generated for each rule (variable length truncation, vowel deletion, rule set) compared with popular abbreviation for each term

	Truncation	Vowel Deletion	Rule	Popular
		Overall, N = 81		
Mean	0.286	0.216	0.283	0.373
Median	0.286	0.143	0.267	0.367
Standard error	0.022	0.023	0.022	0.017
		One Syllable, N = 15		
Mean	0.272	0.519	0.556	0.556
Median	0.233	0.533	0.533	0.533
Standard error	0.053	0.052	0.036	0.036
		Two Syllable, N = 42		
Mean	0.313	0.192	0.208	0.358
Median	0.330	0.173	0.200	0.333
Standard error	0.025	0.024	0.023	0.018
		Three and Four Syllable, N = 14		
Mean	0.408	0.077	0.247	0.317
Median	0.388	0.051	0.224	0.265
Standard error	0.040	0.020	0.056	0.088
		Multiple Words, N = 10		
Mean	0.020	0.024	0.237	0.237
Median	0	0.027	0.173	0.173
Standard error	0.012	0.022	0.038	0.038

deletion set. For two-syllable words all rule set abbreviations were a subset of truncation abbreviations. In the case of multiple words, where an acronym formation rule applied, there was no overlap among the truncation, vowel deletion, and rule set.

Overall, truncation and the rule set were not significantly different from each other, but each was different from the most popular abbreviation (truncation vs. popular, $t_{80} = 3.507$, $p < 0.001$; rule vs. popular, $t_{80} = 5.316$, $p < 0.001$). For one-syllable and multiple words, the rule set was more often the same as people's natural abbreviations than was truncation. However, for two-, three-, and four-syllable words, truncation surpassed the rule set abbreviations.*

It is worth considering the results of allowing variable length truncation and the rule set abbreviations to be used in the same system. If both were allowed for this word set, what proportion of generated abbreviations would be subsumed? In this case the proportion equals 0.415 overall, 0.828 for one-syllable words, 0.313 for two-syllable words,

* In some cases the results of detailed statistical analysis are not presented. We have in all cases given variability estimates for each cell mean. For the most part, the differences among experimental conditions in all experiments were substantial and large relative to the observed variance.

0.408 for three- and four-syllable words, and 0.257 for multiple words. Thus, it is clear that allowing multiple abbreviations accounts for significantly more of the naturally produced abbreviations than any single abbreviation mechanism (truncation vs. combined rules, $t_{80} = 5.512$, $p < 0.001$; rule set vs. combined rules, $t_{80} = 9.846$, $p < 0.001$).

III. EXPERIMENT II: LEARNING RULE-GENERATED VS. POPULAR ABBREVIATIONS

3.1 Introduction

In the second experiment we examined whether learning a set of abbreviations generated by the rules was better than learning those abbreviations most frequently given by the subjects in Experiment I. That is, does placing abbreviations in an internally consistent set result in more memorable abbreviations than people's natural abbreviations?

3.2 Method

3.2.1 Materials

The materials consisted of the 81 command terms and arguments from Experiment I and their abbreviations. There were two abbreviation conditions:

1. Rule Condition: The 81 abbreviations produced by the rule set shown in Table III.
2. Popular Condition: The most frequently given abbreviation for each of the 81 terms.

(Note that for 65 percent of the terms, the abbreviations were identical in the two conditions. That is, for the rule set 65 percent of the terms produced the same abbreviation as the one most frequently given by subjects in Experiment I.)

3.2.2 Subjects

We paid 44 high school students from the Murray Hill, New Jersey, area to participate. There were 23 subjects in the "rule" group and 21 subjects in the "popular" group.

3.2.3 Procedure

Subjects were randomly assigned to either the rule or the popular group. The task was paired-associate learning of the terms and their abbreviations. The rule subjects saw each term paired with the rule-generated abbreviation, whereas the popular group saw each term paired with the most frequently given abbreviation. Subjects in each group were given a randomized deck of 81 cards, each of which contained one term-abbreviation pair. Each subject received a different random order. Subjects studied each card for five seconds, and at the

sound of a buzzer, flipped to the next study card. During an intervening period of 20 minutes, subjects solved logic problems. Then they were given a list with only the terms and were asked to supply the abbreviations they had learned. If unsure, they were to make their best guess.

3.3 Results and discussion

Table V shows the proportion of correct responses for the two experimental groups for all words combined, separately for the abbreviations that were identical for the two groups, and for those that were different for the two groups. It also shows the standard error of the mean calculated across terms. The difference between the rule and popular conditions was highly reliable ($p < 0.001$) in all three cases. Thus, learning an internally consistent set of abbreviations facilitated later recall or generation. (Note that if subjects implicitly or explicitly knew the rules, they should have been able to generate the correct abbreviation for a given term.)

We performed additional analyses to determine in which linguistic environments the rules most facilitated performance. Table VI shows the mean proportion correct as a function of experimental group for one-, two-, three-, and four-syllable words, and multiple words.

The ability of the abbreviation rules to predict the popular abbreviations was far from uniform across the syllable/word classes. For one-syllable words and multiple-word terms, the rule and popular abbreviations were identical. While there was still a recall advantage for the rule abbreviations, it was small and not statistically reliable. However, this was not the case for the other three word classes. Evidently, the abbreviation rules for one-syllable and multiple words are reasonably straightforward and either tacitly known or easily learned by subjects.

The rules and natural productions were in only moderate agreement

Table V—Proportion of abbreviations correctly recalled and/or generated as a function of experimental group

	Rule	Popular
Overall (81 terms)	0.70	0.54
Standard error of the mean	0.02	0.03
Abbreviations same in two groups (53 terms)	0.76	0.63
Standard error	0.03	0.03
Abbreviation different in two groups (28 terms)	0.54	0.37
Standard error	0.04	0.04

Table VI—Proportion correct in rule and popular conditions as a function of the linguistic class of term

	One-Syllable Words N = 15	Two-Syllable Words N = 42	Three-Syllable Words N = 12	Four-Syllable Words N = 2	Multiple Words N = 10
Mean proportion correct					
Rule group	0.78	0.62	0.59	0.50	0.97
Popular group	0.73	0.42	0.45	0.44	0.91
Proportion of items for which rule abbreviation and popular abbreviation were the same	1.00	0.45	0.58	0.50	1.00

for two-, three-, and four-syllable words, and for these word classes, using the rules to produce abbreviations had the greatest facilitative effect on performance. One could presume that there is more variation in terms of linguistic composition for these classes. Owing to this complexity, people's internal abbreviation rules may be inconsistent, incomplete, and therefore, more variable. However, if rules are formulated for subjects, the subjects can use them to learn abbreviations.

3.4 Conclusions

The results and conclusions to be drawn from Experiments I and II are relatively straightforward. First, the process by which we produce abbreviations is not entirely idiosyncratic, but is to a large degree regular. We derived abbreviation principles that characterized a majority of the most frequently elicited abbreviations for each term. The abbreviation rule produced abbreviations that were easier to learn than the most frequently produced abbreviations. There was a substantial advantage of learning abbreviations in a mutually consistent set, even when the abbreviation being learned *was the same* as would be frequently given by subjects. However, the generality of these results is limited, since the rules were derived from the same set of words that served as stimuli in the recall.

IV. EXPERIMENT III: ABBREVIATION PRODUCTION AND LEARNING OF RANDOM WORDS

4.1 Introduction

To test the generalizability of the rule set used in the previous experiment, we applied the rules to a random set of words selected to have the linguistic properties considered relevant in the rule set. A replication was needed, since it is possible that the terms used in the

command set studied in Experiments I and II have properties that differ in some unknown way from ordinary English words. A priori, it is conceivable that the command name words might be different semantically, phonetically, and grammatically from "ordinary" English words. Thus, to know whether rule-based abbreviations are better in general, the test words should be a random sample of words. Also, since the rules were generated from examining subjects' abbreviations of these words, the rule set may include conditions that are the result of idiosyncracies in the particular commands selected.

4.2 Method

4.2.1 Word selection

A total of 200 English nouns and verbs were randomly selected from Kucera and Francis[3]. Words were either one ($N = 40$), two ($N = 80$), or three ($N = 80$) syllables. For the two- and three-syllable words, 40 had prefixes (e.g., "de," "dis," "pro," "post," and "mis"), while 40 did not.

4.2.2 Production subjects

Thirty-two paid Rutgers University undergraduates supplied abbreviations for the random words.

4.2.3 Procedure

The word set was randomly divided in half with half of the subjects supplying abbreviations for 100 words.

4.3 Production results

We analyzed the data in the same way as in Experiment I. Tables VII and VIII show the proportion of abbreviation productions accounted for by variable length truncation, vowel deletion, the rule set, and the most frequently given abbreviation (popular). Table VII shows proportions for all words combined and separately for one-, two-, and three-syllable words. Note that the proportions in Table VII are reasonably close to those found in Experiment I.

Table VIII compares two- and three-syllable words with and without prefixes. The existence of a prefix decreased abbreviation agreement in the popular condition ($t_{79} = 2.790, p < 0.01$). Thus, prefixes affected production. However, it does not appear that prefix conditions in the rule set (as described in Table III) managed to capture any of the differences between words with and without prefixes. Abbreviation behavior was best represented by a truncation rule irrespective of whether the word had a prefix.

Table VII—Abbreviations generated for randomly selected English words

	Truncation	Vowel Deletion	Rule Set	Popular
		All words, N = 200		
Mean	0.396	0.234	0.304	0.406
Median	0.375	0.188	0.312	0.312
Standard error	0.013	0.014	0.015	0.015
		One-Syllable Words, N = 40		
Mean	0.228	0.483	0.473	0.512
Median	0.188	0.500	0.469	0.500
Standard error	0.027	0.033	0.034	0.027
		Two-Syllable Words, N = 80		
Mean	0.382	0.245	0.274	0.374
Median	0.375	0.250	0.281	0.375
Standard error	0.018	0.017	0.019	0.015
		Three-Syllable Words, N = 80		
Mean	0.493	0.099	0.250	0.383
Median	0.500	0.062	0.188	0.375
Standard error	0.019	0.008	0.024	0.017

V. LEARNING RULE-BASED AND POPULAR ABBREVIATIONS FOR RANDOM WORDS

5.1 Method

5.1.1 Subjects

A total of 90 paid adult recruits from the Holmdel area participated in the experiment.

5.1.2 Procedure

The procedure was identical to that reported in Experiment II with the exception that the interpolated task was underlining "important concepts" in a text-editing manual. There were three conditions. Thus, in separate conditions, each word was paired with (1) its rule set abbreviation, (2) its popular abbreviation, and (3) its vowel deletion abbreviation. There were 30 subjects in each of these three conditions.

5.2 Learning results

Table IX shows the proportions and standard errors across words for each of the three groups. In each row each condition was significantly different from every other condition. (The largest matched-paired t probability of occurrence was less than 0.02. However, this probability value is uncorrected, i.e., it does not take into account the number of post comparisons.) Thus, the ordering of generation or recall performance was best for the simple vowel deletion rule, second best for the rule set condition, and worst for the popular abbreviations.

Thus, the original results that rule-governed abbreviations were

Table VIII—Abbreviations generated for randomly selected two- and three-syllable words with and without prefixes

	Truncation	Vowel Deletion	Rule Set	Popular
		Two-Syllable Words, No Prefix (N = 40)		
Mean	0.425	0.259	0.326	0.411
Median	0.438	0.250	0.312	0.375
Standard error	0.028	0.025	0.030	0.021
		Two-Syllable Words, Prefix (N = 40)		
Mean	0.339	0.231	0.222	0.338
Median	0.312	0.188	0.250	0.312
Standard error	0.020	0.024	0.023	0.021
		Three-Syllable Words, No Prefix (N = 40)		
Mean	0.483	0.109	0.308	0.406
Median	0.500	0.094	0.344	0.375
Standard error	0.029	0.012	0.037	0.024
		Three-Syllable Words, Prefix (N = 40)		
Mean	0.503	0.089	0.192	0.359
Median	0.500	0.062	0.125	0.312
Standard error	0.024	0.012	0.029	0.025

better reproduced than abbreviations produced by consensus were strongly replicated. Note that a simple rule produced the best performance in this task. The vowel deletion rule appears to be a particularly easy rule for subjects to abstract. Thus, performance improved as the rules governing abbreviations became more straightforward.

VI. EXPERIMENT IV: RECALLING THE SOURCE WORD GIVEN ITS ABBREVIATION

6.1 Introduction

One of the anecdotal observations frequently made about new computer users is that they learn abbreviations for command terms without ever learning what some of the abbreviations mean. In this sense, the abbreviations represent an alternate name space. There is usually some care given to the choice of names for commands, devices, programs, etc, the assumption being that the use of appropriate names tends to make learning easier (but see Landauer, Galotti, and Hartwell[4], where random words assigned to text-editing functions were no easier or harder to learn or use than appropriate names). However, this care is not usually taken in choosing abbreviations. Most people's goal is to choose abbreviations that are easy to remember, given that the user remembers the right name in the first place. In this experiment we were interested in the decoding process: determining how well the abbreviations represented the words they abbreviated.

There is evidence that the decoding and encoding (determining an abbreviation, given a source word) are asymmetric processes. In par-

Table IX—Mean proportion of abbreviations recalled/produced correctly, given words

	Popular	Rule Set	Vowel Deletion
Overall (200)	0.466	0.650	0.811
Standard error	0.015	0.014	0.006
1-syllable (40)	0.579	0.634	0.875
Standard error	0.029	0.023	0.011
2-syllable (80)	0.428	0.622	0.815
Standard error	0.031	0.035	0.013
3-syllable (80)	0.446	0.685	0.774
Standard error	0.039	0.032	0.011

ticular, rule-based abbreviation schemes are not as effective for decoding purposes as for encoding purposes. With a well-formed rule it is possible to produce the abbreviation accurately when given the word, but the converse is not true. That is, knowing the abbreviation rule and the abbreviation does not necessarily allow one to generate the correct source word.

Rogers and Moeller[5] compared decoding of conventional Navy sonar abbreviations and rule-based abbreviations (produced by truncation). (The subjects in their experiments were sonar operators.) Reaction time to decode the abbreviation was measured. When the scoring criteria were strict, the rule-based abbreviations were worse than the conventional military abbreviations. The difference was due primarily to conventional abbreviations giving a better indication of the endings of words.

Using two 20-word lexicons, Hirsh-Pasek, Niedelman, and Schneider[6] examined decoding in a wide variety of conditions: truncation to four letters, vowel deletion with truncation to four letters, minimum number of letters to distinguish, a phonics system, and user-supplied abbreviations. Vowel deletion and the phonics system produced the fewest number of errors, whereas minimum to distinguish produced the most. Truncation was only slightly worse in terms of errors (perhaps not significantly) than vowel deletion and phonics.

Ehrenreich and Porcu[7] compared decoding of abbreviations formed by vowel deletion and by truncation. There were no reliable differences between these two methods with liberal scoring procedures—65 percent and 57 percent correct, respectively. With stricter scoring criteria, truncation abbreviations were easier to decode (54 percent) than vowel deletion abbreviations (48 percent).

6.2 Method

6.2.1 Subjects

We paid 90 residents from the Holmdel area to participate.

6.2.2 Procedure

We prepared cards using the 200 random English words used in Experiment III. One word and its abbreviation appeared on each card. There were three groups: one group learned popular abbreviations, another the rule set, and a third, the vowel deletion abbreviations. The procedure was the same as in previous experiments, except that after the 20-minute retention interval, subjects were given lists of *abbreviations* and asked to *write down the word* each represented.

6.3 Results and discussion

Table X compares the proportion of words recalled correctly in each of the three groups. Note that in Experiments II and III subjects recalled or produced abbreviations formed by the rule set more often than subjects who learned abbreviations arrived at by consensus. We attributed this improvement in performance to the fact that the existence of the rule lent some organization to the set of stimuli being learned. When we looked at ability to recall what the abbreviations represent, however, it appeared that the rule set seemed to interfere with subjects' ability to recall what the abbreviations stood for. Thus, there is an asymmetry between recalling the abbreviation from seeing the command term and recalling the command term from seeing its abbreviation. The superiority of the vowel deletion rule in this task is evident for all but the one-syllable words, and for this case the rule set is essentially a vowel deletion rule.

These results are somewhat at odds with those of Ehrenreich and Porcu.[7] First, in our experiment, the vowel deletion condition produced very high performance. Also, the overall level of performance was much higher in our study (79 percent here vs. 48 percent in the Ehrenreich and Porcu study). Ehrenreich and Porcu claim that when subjects produce words from vowel deletion abbreviations they are apt

Table X—Mean proportion of words recalled/produced correctly, given abbreviations

	Popular	Rule Set	Vowel Deletion
Overall (200)	0.626	0.467	0.788
Standard error	0.022	0.024	0.018
1-syllable (40)	0.662	0.635	0.641
Standard error	0.046	0.050	0.052
2-syllable (80)	0.628	0.500	0.799
Standard error	0.037	0.038	0.027
3-syllable (80)	0.605	0.350	0.850
Standard error	0.035	0.034	0.024

to make spelling errors. People who misspell a word are more likely to err on a vowel than a consonant. Their terms were Army terms—ours random words; their subjects were Army personnel—ours primarily spouses of Bell Laboratories employees. These differences may have produced the resulting performance differences.

It would seem that vowel deletion should have produced the best results, since consonants carry more information than vowels and the abbreviations in this condition were longer than in the other two conditions (vowel deletion = 4.43 letters, standard deviation = 1.13; rule condition = 3.38 letters, standard deviation = 0.62; popular = 3.94 letters, standard deviation = 1.08). Thus, the results vary directly with the number of letters in the abbreviation alone.

In summary, there was a large asymmetry between the encoding results for the popular and rule condition in Experiment III and the decoding results in Experiment IV. Rule abbreviations were better than popular for remembering the abbreviation given the word, whereas popular abbreviations were better than rules for remembering the word given the abbreviation.

VII. GENERAL DISCUSSION

While there is much variability in people's natural abbreviations, the process is far from random. People are most likely to abbreviate one-syllable words by deleting word-internal vowels, and multiword terms by deleting everything except the first letter of each word. Abbreviations of polysyllabic words are best characterized by truncation, that is, deleting letters from the end of the word. While these rules accounted for the majority of the most frequently generated abbreviations for each term, the proportion of all abbreviations covered by these rules was only about 0.30 in the two abbreviation production studies. Thus, there is much that people do when abbreviating that is idiosyncratic, variable, and not particularly amenable to rule-based descriptions.

We showed that if abbreviations are formed by applying rules derived from people's behavior, subjects can use these principles. Thus, subjects make use of the internal consistency of the set to learn the abbreviations. Performance with a set of rule-based abbreviations was better than with the corresponding set of most frequently produced abbreviations even when the two abbreviations were identical in the two conditions.

That people extract principles in complex stimuli is not a new finding. For instance Reber,[8,9,10] among others, has demonstrated that people's behavior when learning artificial languages in an experimental setting indicates that they have "learned" the grammatical rules of the language. However, very often they are unable to verbalize these rules

and seem unaware that they have extracted principles. It appears that rule formulation is often an automatic and unintentional process.

Throughout, we have explicitly assumed that rules based on behavior are better than arbitrary rules for abbreviation. One could argue that our own data contradict this assumption. That is, the vowel deletion rule used in Experiments III and IV produced the best performance in two tasks—recalling the abbreviation given the word and conversely recalling the word given the abbreviation. Don't the data then argue that one should abbreviate using vowel deletion? Yes, if the abbreviation is used to recall the term, that is, a decoding task. However, the ability to construct a correct abbreviation, given the word, is only one aspect of a good abbreviation. Another aspect that "abbreviate" itself denotes is that it should be short. Abbreviations produced by vowel deletion increase in length as the length of the source word increases, whereas abbreviations that people spontaneously produce and the ones produced by the rule set better maintain length constancy [e.g., a one-syllable word on the average is abbreviated to 3.4 letters in all conditions, whereas a three-syllable word is abbreviated to 5.2 letters (vowel deletion), 4.3 letters (popular), and 3.3 letters (rule)]. Thus, in terms of total number of keystrokes, the simple vowel deletion rule is least efficient.

A more serious failing for simple vowel deletion is that forming the abbreviation requires first producing the whole word, then systematically deleting the vowels, and finally outputting the remaining consonants. We suspect that this is difficult for long words. When a word is short, the entire word can be dealt with as a "chunk." For a single chunk it is probably not difficult to remove a vowel or two. Beyond simple, short words, the amount of mental bookkeeping increases. If one's goal is to enter command strings *quickly*, vowel deletion is probably not a serious candidate.

Truncation, on the other hand, has properties that make it the best abbreviation mechanism for polysyllabic words. Resulting abbreviations are short and easy to produce—the user does not need to pervert the assumed normal output strategies, just cut them short. It is a simple rule to teach[5–7] and users are most likely to abbreviate multisyllabic words in this way. Thus, it seems to have most of the desirable properties of an ideal abbreviation mechanism—natural, short abbreviations, easy to remember abbreviations, and easy to produce abbreviations. The one thing that may be sacrificed is uniqueness.

There have been suggestions on how to maintain uniqueness of abbreviations, which we now discuss. Many of the recommendations require the user to be flexible and relearn abbreviations from time to time, and some recommendations undermine the consistency of a rule-based set. Ehrenreich and Porcu[7] discuss two ways to resolve conflicts

among abbreviations: (1) use of alternate rules and (2) minimum number of characters to distinguish.

Using an alternate rule when conflicts exist has some serious problems. First, the user must know that there are two or more rules, and that one words most of the time, but when it doesn't, to use the alternate rules. Furthermore, the user has to know which one of the conflicting words requires which rule. The user doesn't know in advance when the primary rule will not work and, consequently, will probably have to learn by trial and error. When the number of exceptions to the primary rule is large, we suspect that the advantages gained by selecting abbreviations by rule will disappear. That is, the mixed rule set performance may approach or may even be worse than performance on abbreviations selected by consensus. However, the degree to which rule-based systems tolerate corruption warrants investigation.

In the case of minimum to distinguish, the user truncates to the number of characters required to differentiate that word for all other admissible words. For example, for "transport" and "transfer," the minimum to distinguish is in each case six characters—"transp" and "transf." There are a few problems worth noting with respect to minimum to distinguish. First, different words require different numbers of letters. Hirsch-Pasek et al.[8] have found in a paired-associates task that minimum to distinguish was more difficult to learn than the other abbreviation schemes they studied; truncation to four letters was the easiest. This may be only a relatively minor problem, which depends on the terms of the command set and the size of the lexicon. It could be dealt with by telling the user the minimum number of letters needed to distinguish all command names. However, a fixed minimum based on all commands will be unnecessarily long for many commands. The user could of course be told that fewer keystrokes will often suffice and learn the minimum to distinguish on a command-by-command basis. The second potential shortcoming occurs when new commands that produce collisions are introduced into the language. Truncations that worked previously are now ambiguous and require relearning. However, introducing software that informs the user of a conflict could solve these problems. If the input is ambiguous, the system could present the user with the collisions and ask the user to select one of the alternatives.

One way to minimize the likelihood of collisions is to make contextual information available to the system. For example, there may exist three words—"debug," "delete," and "define." If collisions were considered on the basis of the entire lexicon, three letters would be needed to differentiate these words. However, if "define" were a command while "debug" and "delete" were command options that occurred in

different environments, it could be the case that a "d" would suffice for either "debug" or "delete." Thus, if the command/option tables are structured hierarchically, such that a command points only to options meaningful in the context of that command, the potential range of conflicts is reduced from the entire command language to a small subset of the language.

In summary, considering all available evidence on producing abbreviations given command terms (encoding), truncation appears to be the best single abbreviation scheme. Truncation also best captures people's natural abbreviations in all environments except two—monosyllablic words and multiple-word terms. In these cases, we recommend using vowel deletion for the former and acronym formation for the latter. If, on the other hand, one's task requires generating full names, given abbreviations (decoding), vowel deletion abbreviations are better than other rule-based schemes.

REFERENCES

1. G. Zipf, *The Psycho-biology of Language*, Boston: Houghton Mifflin, 1935.
2. M. Hodge and F. M. Pennington, "Some Studies of Word Abbreviation Behavior," J. Exp. Psychol., *98*, No. 2 (1973), pp 350–61.
3. H. Kucera and W. N. Francis, *Computational Analysis of Present-Day American English*, Providence, Rhode Island: Brown University Press, 1967.
4. T. K. Landauer, K. M. Galotti, and S. Hartwell, "Natural command names and initial learning: A study of text editing terms," Commun. ACM, *26*, No. 7 (July 1983).
5. W. H. Rogers and G. Moeller, "Evaluation of Abbreviations in Sonar Displays," 52nd Annual Meeting of the Eastern Psychological Association, New York City, April 24, 1981.
6. K. Hirsh-Pasek, S. Niedelman, and M. L. Schneider, private communication.
7. S. L. Ehrenreich and T. A. Porcu, "Abbreviations for Automated Systems: Teaching Operators the Rules," *Directions in Human Computer Interaction*, A. Badre and B. Shneiderman (Eds), Norwood, NJ: Ablex Publishing Company, 1982, pp. 111–135.
8. A. S. Reber, "Implicit Learning of Artificial Grammars," J Verb. Lear. Verb. Behav., *6* (1967), pp. 855–63.
9. A. S. Reber, "Transfer of Syntactic Structure in Synthetic Languages," J. Exp. Psychol., *81*, No. 1 (1969), pp. 115–9.
10. Reber, A. S., "Implicit Learning of Synthetic Language: The Role of Instructional Set," J. Exp. Psych., Human Learning and Memory, *2*, No. 1 (1976), pp. 88–94.

AUTHORS

Lynn A. Streeter, B.A. (Psychology), 1969, University of Michigan; Ph.D. (Psychology), 1974, Columbia University; Bell Laboratories, 1969—. Ms. Streeter first joined Bell Laboratories in 1969 as a member of the then Human Factors Department in Holmdel. She rejoined Bell Laboratories in 1974 as a member of the Linguistics and Speech Analysis Department. The abbreviation data were collected when she was Supervisor of the Human Factors Engineering Group for the Advanced Communications Service project. Currently, she is working on mobile telephony interface design problems in the Communications Methods Research Department.

John M. Ackroff, B.A. (Psychology), 1971, Williams College; M.S. (Experimental Psychology) 1975, Ph.D. (Experimental Psychology) 1981, University

of Wisconsin–Milwaukee; Bell Laboratories, 1977–1982; American Bell, 1982—. While at Bell Laboratories, Mr. Ackroff worked on the Advanced Communications Service project, where he was responsible for the design and specification of the command language aspects of the user interface. At American Bell, he is responsible for developing an enhanced version of the command language parser.

Glen A. Taylor, B.S. (Psychology), 1969, Louisiana State University; M.A. (Experimental Psychology), 1975, Ph.D. (Experimental Psychology), 1977, University of Kansas; Bell Laboratories, 1979–1982; American Bell, 1982—. Mr. Taylor joined Bell Laboratories as a member of the Human Factors Engineering Group for the Advanced Communications Service project (now AIS/Net 1000 service). He was responsible for designing the user interface for the Net 1000 editor and for investigating various aspects of computerized electronic mail services. In 1981, he became a member of the User/Station Interface Software Group responsible for the development of on-line documentation retrieval software.