INSTITUTE OF COGNITIVE SCIENCE

Technical Report

University of Colorado, Boulder

Acquisition and Retention of Morse Code Reception Skills

by

Deborah Marie Clawson

Institute of Cognitive Science
University of Colorado
Boulder, Colorado 80309-0344

ICS Technical Report #95-04

Acquisition And Retention Of Morse Code Reception Skills

Deborah Marie Clawson

Institute of Cognitive Science
University of Colorado
Boulder, Colorado 80309-0344

ACKNOWLEDGEMENTS

This paper is based on my master's thesis. I wish to thank my advisor, Alice F. Healy, and committee members Lyle E. Bourne Jr. and K. Anders Ericsson. I am grateful to Bob Wisher and to Stephanie Doane for helpful discussions and for initially stimulating the interest in this project. Thanks also go to Charles M. Judd and Gary H. McClelland for their statistical advice. This research was supported by the Army Research Institute under Contract MDA903-90-K-0066.

Abstract

Several aspects of receiving singly presented Morse codes were investigated. Pellegrino, Doane, Fischer, and Alderton's (1991) finding of an acquisition advantage for initial training on the most difficult subset of a task's stimuli rather than on its easy subset was found not to apply to the task of Morse code reception, nor was there a retention advantage for difficult initial training. Examination of two subtasks underlying the task of Morse code reception revealed that the pattern of errors made on the first subtask and on the whole task supported previous research suggesting that novice knowledge of single Morse codes is organized by number of elements in the codes. The subtask of recognizing the pattern of dots/dashes in the code is a relatively stable skill that shows no forgetting over retention, whereas the subtask of accessing the arbitrary letter corresponding to that code does show decline over retention. These retention results were consistent with the procedural reinstatement theory of skill retention. However, operational definitions of procedural and declarative memory as they relate to procedural reinstatement must be developed.

CONTENTS

CHAPTER

TABLES

Table

FIGURES

Figure

CHAPTER I

INTRODUCTION

Three areas of research are particularly relevant to the study of acquisition and retention of Morse code skills. The first area is that of research on Morse code skills themselves. The second area is that of research on part-whole training. The final area is that of the implications of procedural reinstatement for retention performance.

Morse code reception

Psychological research in Morse code reception is near celebrating its centennial. By 1899 Bryan and Harter had already published a review of their previous research on the acquisition of proficiency in sending and receiving Morse code by telegraph operators. Bryan and Harter examined the learning curves for meaningful text over a period of many months, concluding that speed-up in performance was first due to a speed-up in reception at the letter level, then to speed-up in receiving words as units, and finally to higher-order language processing of the codes at the sentence level.

More recent research on Morse code skills has included descriptions of the component skills of Morse code based on clinical evidence from Morse code aphasia, a deficit in communicating with Morse code (Ardila, 1987; Wyler and Ray, 1986). Ardila argued from personal observation that Morse code skills involve three components: an auditory perceptual component, a motor sequencing component (for sending code), and a linguistic component. However neither of the case studies explicitly showed dissociations among these components, so the neuropsychological data cannot support the component description.

Training methods for Morse code have been examined, yielding such conclusions as that spreading out in the practice order presentation of similar codes leads to higher accuracy than does grouping the similar codes (Rothkopf, 1958). Newby, Cook, and Merrill (1988), investigating visual mediation in the training of Morse reception, found that visual mediators led to higher final acquisition levels and retention levels than no mediators. Campbell and Allen (1987) gave subjects one half-hour of instructor-paced or self-paced training and found that the training condition did not significantly affect test accuracy.

One real-life study of the retention of Morse code skills involved Morse code sending skills rather than reception skills; it found that recruits experiencing delays of up to ten weeks between training and arrival at their assigned duty stations exhibited error rate increases of 1.2 percent or less (Morsh & Stannard, 1947).

Training using a whole method, wherein subjects learn all the letters simultaneously, can be effective. Keller (1943) demonstrated that 81 percent of his subjects reached criterion (three successive 36-item blocks, each with 95 percent correct) after twelve hours practice on single-character Morse code reception on a whole-set of all 26 letters and 10 numbers. Spragg (1943), using Keller's instruction method, demonstrated that 65 percent of his subjects reached criterion performance (at least 95 percent correct on a single 100-item block) after nine 300-trial sessions of practice on a whole-set of the 26 letters. Other studies during World War II included many studies of the errors made in Morse code reception; these studies are discussed in Chapter IV.

Klapp and Wyatt (1976) found that subjects using a telegraph key to respond with one of the 2-element codes (. . , . _ , _ . , or _ _) showed that initial reaction times were longer for mixed-element codes than for homogeneous codes. Also, the subjects depressed the key longer during the first element and displayed a longer intra-character gap if the ensuing element was a dash than if it was a dot; Klapp and Wyatt suggested that this indicated motor programming for the second element took place during the first element and the intra-character gap rather than preceding response initiation.

Cormier, Tomlinson-Keasey,and Geary (1988), using dot-dash visual presentation of codes, showed that on a verification test women showed quicker correct reaction times than men. Cormier et al. also claimed that cerebral hemispheric specialization for the codes (as indicated by visual field advantages in accuracy) follows a different pattern in men than in women, perhaps reflecting more lateralized language function in men than in women; no hemispheric advantages were found on reaction time measures.

Part-whole training

There are two main systems for predicting the efficacy of whole versus part training. Naylor and Briggs (1963) proposed that the effectiveness of progressive part versus whole training depends on the task (in progressive part training, one first practices subtask one, then practices subtasks one and two together, then practices subtasks one through three together, and so on). They held that, for a highly integrated task, whole training is always better; for a loosely integrated task whose subtasks are independent, progressive part training becomes more effective than whole training as the complexity of the subtasks increases. However, Naylor and Brigg's designation of their tasks as integrated or independent is only relative, as is the designation "complex"; thus the scheme cannot easily be applied to new tasks.

The second system of predicting the success of part training is that of Annett and Kay (1956). Annett and Kay proposed that whole training is best for skills in which the responses on early parts of the task do not influence success on later parts of the task. For tasks in which responses in early parts of the task influence later parts of the task -- such as driving a car -- part training should be used to eliminate errors in the early parts of the task. Their recommendations can be applied to the task of Morse code reception. Receiving a Morse code is a task in which errors on the first part of the task, recognizing the pattern of dots and dashes, can render impossible the correct performance of the second part, identifying that pattern as its corresponding letter. If the subject segments the code incorrectly, then in whole-task training that subject cannot learn the code-letter correspondences because the same letter may appear to be associated with many different codes and vice versa.

Stammers (1982) in a review of part-whole literature recommended that in general one should begin by using whole training; only if whole training does not work in a pilot training session, then switch to part should part training be considered. Besides other more practical considerations, the basis of Stammers's recommendation was his conclusion that research has not often found an advantage for part training.

Studies finding advantages for part training are not unheard of, however. Mane, Adams and Donchin (1989) found that 14 minutes of prior part training led to higher performance as measured by score on a video game throughout 20 blocks of practice. The prior part training consisted of two cumulative part drills -- that is, the second practice task subsumed the initially practiced task -- followed by a third drill that was not cumulative. Seymour (1954) suggested that when a task contains a mixture of difficult and easy parts learning the difficult parts first leads to better performance.

Newell, Carlton, Fisher and Rutter (1989) discussed an important issue in part training. They insist that part training should use "natural" subtasks. Training on a "natural unit of coordinated activity" was more effective than training that focused isolated components of that coordinated subtask. Newell et al. acknowledged, however, that the operational definition of "natural" subtask is not clear.

Procedural reinstatement

Kolers and Roediger in 1984 presented evidence for a procedural account of learning and memory, proposing that the procedures that characterize a person's acquisition of knowledge are important, that "the effects of experiences depend upon the procedures used to realize them" (p. 436). Healy , Fendrich, and Proctor (1990), in considering the procedural account of memory put forward by Kolers and Roediger, proposed that learning procedures, inextricably linked to the learned skill, have crucial effects on skill retention. Specifically, skills with procedures that are easily reinstated at test will show a substantial degree of retention -- in some cases perfect retention -- whereas those that do not easily prompt this procedural reinstatement, either because a different procedure is required at retention than at acquisition or because the learning procedure cannot be recalled at retention, will show greater forgetting (Healy, et al., 1990; Healy et al., in press).

Anderson in 1976 described an important dichotomy, the dichotomy of procedural and declarative memory. Tulving (1985) also distinguishes between procedural and other types of memory (while describing as well the distinction between semantic and episodic memory), and Squire (1987) gives evidence from amnesia for two separate kinds of memory. Essentially, procedural knowledge is "knowing how," and declarative knowledge is "knowing that"; procedural knowledge prescribes how to do something, and declarative knowledge describes the world (Anderson, 1976; Tulving, 1985). Three features of declarative (or, for Tulving, semantic) versus procedural information are: procedural knowledge can be partially possessed whereas declarative knowledge is all-or-none; procedural knowledge is acquired gradually by performing the task, whereas declarative knowledge can be acquired suddenly by simply being told; and procedural knowledge can be communicated only by direct expression, by doing it, whereas declarative knowledge can be communicated flexibly, to include verbally (Anderson, 1976; Squire, 1987; Tulving, 1985). Squire adds that procedural memory can be modality-bound, whereas declarative memory is not. Anderson's examples of procedural skills include riding a bicycle, speaking one's native language, and driving a car; examples of declarative information are recalling the fact that Washington was the first president of the United States, speaking a new foreign language, and initially driving a stick shift car. As the driving examples imply, the line between procedural and declarative information is not always easy to draw, and Anderson argues that with practice there can be a shift from declarative to procedural memory.

The importance of the distinction between procedural and declarative tasks to procedural reinstatement is that procedural skills explicitly require the use of the same procedure at retention as during training. For declarative information the learning procedures are more indirect, and thus declarative information will fare more poorly at retention unless it is acquired using a readily reinstated mnemonic technique(Healy et al., in press).

CHAPTER II

EXPERIMENT 1: THE EFFECT OF INITIAL TRAINING SUBSET DIFFICULTY ON ACCURACY FOR MORSE CODE RECEPTION

When teaching Morse code, is it more effective to begin by teaching the easy letters or to begin by teaching the difficult letters? Such a question is basic to any subject wherein the items to be learned vary in difficulty. A recent visual discrimination study by Pellegrino, Doane, Fischer, and Alderton (1991) suggests that the most effective training starts with the more difficult stimuli. In that study subjects who first learned the more difficult discriminations performed better than subjects trained first on the easy discriminations, after both groups had been given further practice on the full set of stimuli.

This result cannot be simply generalized to the task of Morse code reception because it differs from Pellegrino et al.'s task in three important ways. First there is the difference in modality, because Morse code reception is an auditory task. Second, whereas the Pellegrino et al. task was discrimination of stimuli that were both present at the time of response, Morse code reception calls for identification, in essence discriminating the presented stimulus from all other (not present) stimuli in the set plus labelling that particular stimulus. Third, Pellegrino et al.'s task allowed high accuracy, so the advantage in performance was found only on reaction time. In the early stages of Morse code reception training, accuracy is not near ceiling and therefore remains the chief concern. It remains to be seen whether the advantage for practice starting with more difficult stimuli would be reflected in improved accuracy.

Pellegrino et al. propose as a cause of the advantage that the subjects trained on easy stimuli acquired a very loose net of discriminations. Because the discriminations were easy, a loose net was enough; they only needed a general strategy for comparing two presented stimuli to see whether they were different. The subjects trained first on hard discriminations were confronted with stimuli that were not so easily conquered using this general comparison strategy. These stimuli, Pellegrino et al. argue, were challenging enough to encourage subjects to learn to identify each of the stimuli, to cast a much finer net over them. They adopted stimuli-specific strategies. The greater specificity of their stimulus encoding network allowed them better performance.

The implications of this proposal are clear for Morse code reception: Morse codes can be loosely distinguished on the basis of general strategies or tightly distinguished using item-specific strategies. The codes can be distinguished generally because they vary across a number of dimensions: length in time, degree of homogeneity of elements (dots and dashes), total number of elements, number of a single element type. The letters can also be distinguished using item-specific strategies: A letter can be identified by its element pattern, by its rhythmic similarity to words or phrases, by images or motions that it calls to mind, by processes of elimination from the codes of which the subject is sure.

In testing the applicability of Pellegrino et al.'s findings to the task of Morse code reception, this study presented subjects with a set of Morse stimuli consisting of the full complement of two- and three-element codes. The set was divided into those codes often confused with others in the set, and those codes less often confused with those in the set, with the restriction that each set contained two 2-element codes and four 3--element codes. The codes were divided into these two confusability sets after reference to the tables of error tallies presented by Keller and Taubman (1943) for novice learners of all 36 letters and numbers of Morse code. In Experiment 1, one group of subjects (the easy-first group) learned first the six easy codes, another group (the difficult-first group) learned first the six difficult codes, and the final group (the all-first group) learned first the complete set of twelve codes. All subjects were then trained further on the complete set of codes. In Experiment 2, there were only two groups of subjects, those trained first on the easy codes and those trained first on the difficult codes; again all subjects were then trained further on the complete set of codes. If Pellegrino et al.'s findings can be applied to the Morse code reception task, then the subjects initially trained on the difficult codes should perform better after further training on all codes than would those initially trained on the easy codes.

The predictions based on Pellegrino et al.'s findings apply only to acquisition not to retention of a skill, because their study did not look at retention. Experiments 1 and 2 therefore further examined the effect of initial training subset on retention with a test administered four weeks after the last day of training.

One further hypothesis was briefly examined. In a review of the skill retention literature Farr (1987) concluded that the most important factor predicting retention level is final acquisition level. Therefore, correlations between final acquisition performance and retention performance were collected to determine the predictive validities for this task.

Method

Design. The design used was a 3 (initial training condition) x 2 (test order) x 4 (session) x 2 (pretest/posttest) x 2 (code-letter difficulty) mixed factorial design. The first two factors, initial training condition and test order, were varied between subjects. The remaining factors -- session, pretest/posttest, and code pair difficulty -- were varied within subjects.

Subjects. The subjects were students in introductory psychology classes at the University of Colorado at Boulder who participated in order to fulfill a class requirement. All were native English speakers who did not have previous knowledge of Morse code. There were four subjects per initial training condition, with two in each condition assigned to each of the two test orders. Assignments were made on the basis of a fixed rotation according to time of arrival for the first session.

Materials. The codes were presented aurally by an IBM PC, and subjects typed their responses using the PC's keyboard. The keys for the twelve letters in the stimulus set were marked with black tape to distinguish them from the letters not to be used.

The tests each consisted of four blocks of the full set of codes, each block randomly arranged.There were eight tests, a different one for each test time. All subjects were presented with all of the tests, but the tests were presented in opposite orders for the two order conditions -- that is, in Order 1 the tests were presented in sequence from one to eight; in Order 2 they were presented in reverse sequence from eight to one.

Procedure. The participants were initially trained under one of three conditions. The easy-first group of subjects was initially trained (that is, trained during the first session) on only the easy half of the stimulus set, the code-letter pairs for I, M, O, S, R, and K. The difficult-first group was initially trained on only the difficult half of the stimulus set, the code-letter pairs for A, N, G, U, W, and D. The all-first group was initially trained on the full set of 12 code-letter pairs. During the second and third sessions, two and four days later, all subjects received training on the full set of stimuli. The 12 code-letter pairs in the stimulus set make up the complete set of two- and three- element codes (see Figure 1).

In the initial session of training, the subjects first were given four blocks of typing practice, with each block consisting of the twelve letters in the response set, randomly arranged. On each trial, the computer displayed one of the 12 letters on the screen with the words "Type this letter" and the subject responded by typing a key. After each trial, the computer responded with accuracy feedback ("Correct response!" or "Wrong response!") and, if the response was correct, with the initial reaction time in seconds. After a two-block typing test on those twelve letters (without feedback), the subjects then began learning the Morse codes. First, the computer cycled through the full set of twelve pairs four times, playing the code

Easy code-letter pairs:

I	. .	S	. . .
M	_ _	O	_ _ _
		R	. _ .
		K	_ . _

Difficult code-letter pairs:

A	. _	D	_ . .
N	_ .	G	_ _ .
		U	. . _
		W	. _ _

Figure 1. The full set of 2- and 3-element codes

while showing the letter on the screen. After listening to these stimuli, the subjects were presented with a 48-item pretest covering all the code-letter pairs four times with the pairs of each block presented in random order. After the pretest the subjects went through 120 trials of practice on their initial stimulus subset; this practice was arranged in 20 blocks of the randomly presented six pairs.The first session ended with another 48-item test, the posttest.

The second and third training sessions each consisted of a 48-item pretest; 120 trials of practice, this time ten blocks of all 12 code-letter pairs; and a 48-item posttest. The Retention Session, four weeks after the third session of training, was identical to the second and third days of training. It consisted of a 48-trial pretest, 120 trials of practice on the full set of letters, and a 48-trial posttest.

The procedure for a training trial was as follows. First the computer displayed the phrase "Type the letter for this code" and played a Morse code out loud; then it prompted the subject ("Letter:"). After the subject typed a letter, the computer gave feedback. If the subject was correct, the computer responded with "CORRECT" and the initial reaction time in seconds; if the subject was incorrect, the computer responded with "INCORRECT" and displayed the correct letter (in the phrase "That was the code for ..."). Finally, regardless of the subject's accuracy, the computer reviewed the trial by playing again the code and displaying the correct letter.

The test trial procedure was identical to the training trial procedure except that there was no feedback and no review.

Four of the subjects, two from the difficult-first group and two from the easy-first group, were trained on giving talk-aloud and retrospective verbal reports. During each session, they were asked for both types of verbal report during approximately one sixth of the training trials and asked for only retrospective reports after each occurrence of four specific codes on all tests as well as after all 12 codes on the final block of the posttest. Information from the retrospective reports will not be reported here.

Results and Discussion

All analyses of accuracy performance were completed using individuals' arcsines of the square root of the proportions correct, because many of the proportions fell outside the range 0.2 - 0.8. All means provided are anti-arcsine; that is, they are the squared sines of the arcsine means. Reaction times were not analyzed because retrospective reports were collected from some of the subjects.

Performance during acquisition. None of the initial training conditions lent a lasting advantage during acquisition, although initial training did lend each part-set group an immediate temporary advantage on its practiced set. Performance during acquisition confirmed the a prior classification of subset difficulty.

The pattern of accuracy performance over training and retention for subjects in the three initial training conditions is illustrated in Figures 2 and 3. A five-way mixed analysis of variance (ANOVA) was conducted on accuracy during acquisition, the first three sessions (see Table 1). The main effect of initial training condition was not significant, $F(2, 6) < 1$; nor was the main effect of test order, $F(2, 6) < 1$. Improvement in performance due to practice was evident, with the main effects of session and of pretest/posttest both significant. Performance on the difficult code-letter pairs was significantly poorer than on the easy pairs, supporting the a priori classification of the stimulus subsets.

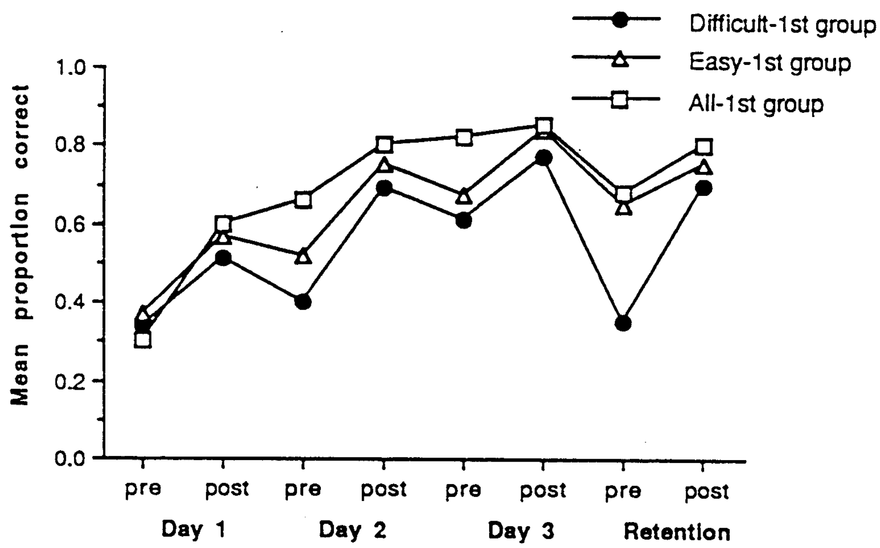


Figure 2. Experiment 1 mean accuracy over acquisition and retention sessions

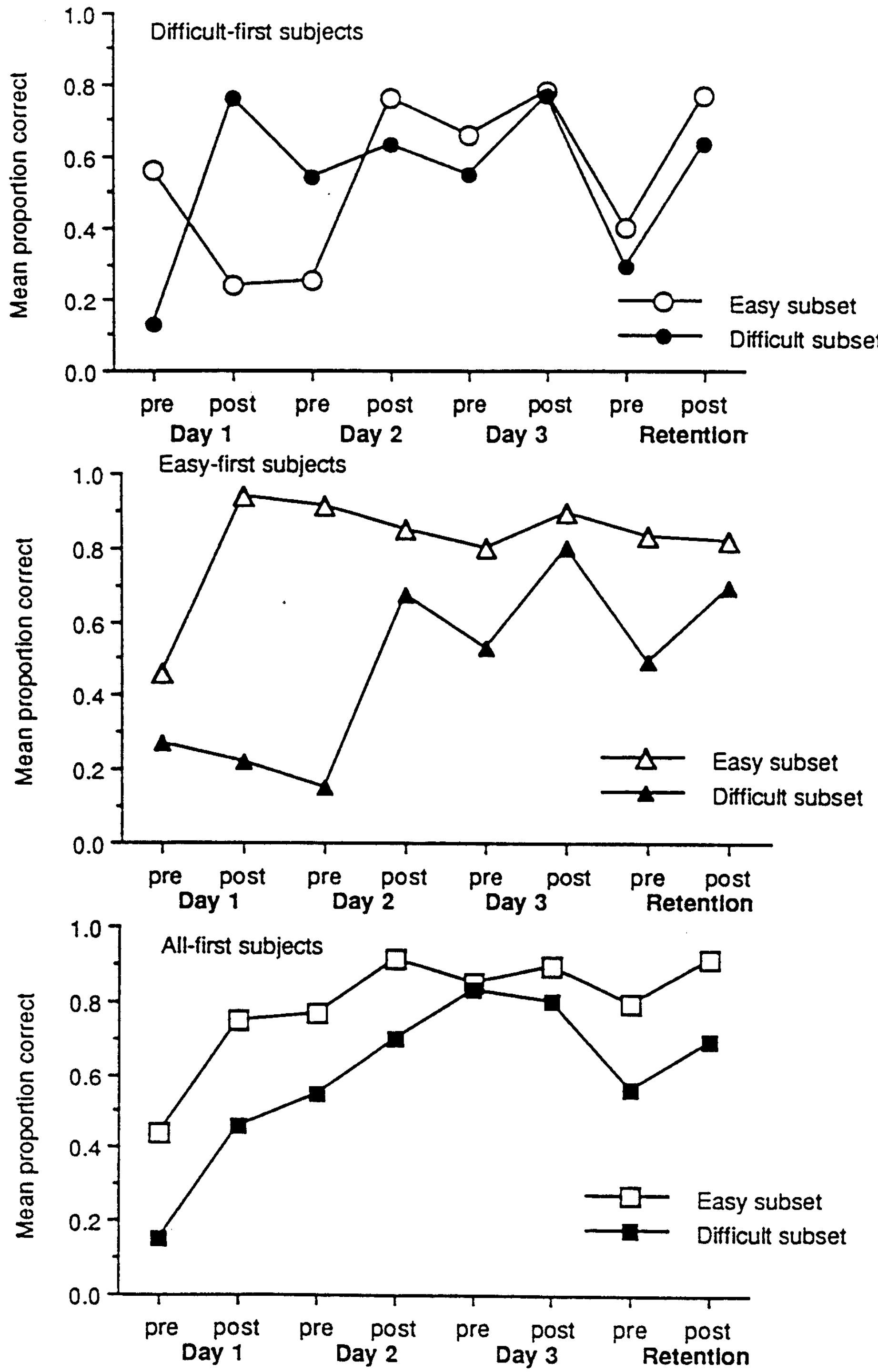


Figure 3. Experiment 1 mean accuracy over training and retention on easy and difficult code-letter pairs

Table 1

Experiment 1 Arcsine Proportion Correct ANOVA Results for All Effects with *p* < .05 on Acquisition

Variable	df	F	p	MSe
Session	2, 12	40.1	< .001	.035
Pretest/posttest	1, 6	176.9	< .001	.009
Letter difficulty	1, 6	35.8	.001	.043
Initial training condition x letter difficulty	2, 6	13.7	.006	.043
Test order x session x letter difficulty	2, 12	5.1	.024	.025
Initial training x session x pre/post x ltr difficulty	4, 12	23.7	< .001	.021

Directly after initial training the difficult-first group exhibited much higher accuracy on the difficult subset than on the easy subset; the easy-first group exhibited much higher accuracy on the easy subset than on the difficult subset. This temporary advantage on their practiced subsets for the two part training groups after initial training (i.e., Session 1 posttest and Session 2 pretest) is reflected in the significant interactions of initial training condition and letter difficulty and of initial training condition, session, pretest/posttest and code-letter difficulty. In post hoc analyses (with a Scheffe adjustment using m = 5) of accuracy on the Session 1 posttest and the Session 2 pretest, the interaction between training condition and letter difficulty was significant on the Session 1 posttest ($F(2, 8) = 44.0$, $p < .005$, MSe = .021) but not for the Session 2 pretest ($F(2, 8) = 13.1$, $p > .070$, MSe = .046) .

The interaction between test order, session and letter difficulty was significant. Subjects tested in Order 2 showed higher accuracy than those in Order 1 on the easy subset of Session 1 tests and the difficult subset of Sessions 2 and 3 tests (see Table 2). Because the effect was not influenced by initial training condition, it is not theoretically interesting.

Table 2

Experiment 1 Mean Proportions Correct (anti-arcsine) during Acquisition

	Easy Subset Order 1	Easy Subset Order 2	Difficult Subset Order 1	Difficult Subset Order 2
Session 1	.50	.66	.32	.31
Session 2	.78	.74	.44	.63
Session 3	.83	.81	.68	.76

A planned analysis of accuracy at the Session 3 posttest showed that the effect of letter difficulty remained significant, $F(1, 8) = 5.8$, $p = .043$, $MSe = .011$, but initial training condition did not significantly affect this final level of acquisition, $F(2, 8) < 1$. Final acquisition level also did not show a significant influence of training condition on the effect of letter difficulty as it had immediately after initial training, $F(2, 8) = 1.9$, $p = .217$, $MSe = .011$.

Performance at retention. Although there was a trend suggesting a Retention pretest disadvantage for the difficult-first group, this disadvantage was not significant in a planned analysis of that test alone. Performance at retention reflected significant forgetting and again confirmed the a prior classification of subset difficulty.

To examine the role of initial training set on retention accuracy, a five-way mixed ANOVA was conducted on accuracy for only the last two sessions, Session 3 and the Retention Session (see Table 3). One subject did not return for the Retention Session; that subject's missing values were replaced by the scores of the other subject in that cell (i. e., in the same initial training condition and same test order) for this ANOVA.The main effect of initial training condition was not significant, $F(2, 6) < 1$; nor was the main effect of test order, $F(1, 6) < 1$. As would be expected, test scores at retention were lower than those at Session 3, and scores were higher on the posttests than on the pretests. As in acquisition, accuracy was lower for the difficult subset than for the easy subset.

Table 3

Experiment 1 Arcsine Proportion Correct ANOVA Results for All Effects with $p < .05$ on Retention

Variable	df	F	p	MSe
Session	1, 6	13.7	.010	.023
Pretest/posttest	1, 6	47.4	< .001	.015
Letter difficulty	1, 6	17.3	.006	.040
Initial training condition x pretest/posttest	2, 6	5.5	.043	.015
Session x letter difficulty	1, 6	8.4	.027	.008

The significant interaction between session and letter difficulty reflects the greater drop in performance on difficult code-letter pairs than on easy pairs between Session 3 and the Retention Session. The interaction between training condition and pretest/posttest was significant; the difficult-first group scored lower on the pretests relative to the posttests than did the other two groups.

The difficult-first group scored lower on the Retention pretest than did the other two groups, suggesting that forgetting was greatest for the difficult-first group. This effect of initial training condition, however, was not significant in a planned analysis of that test, $F(2, 8) = 2.0$, $p = .200$, $MSe = .128$; nor was the interaction of initial training condition and letter difficulty, $F(2, 8) = 1.5$, $p = .275$, $MSe = .018$; the effect of letter difficulty remained significant, $F(1, 8) = 8.6$, $p = .019$, $MSe = .018$.

Predictive validity. Across initial training conditions, final acquisition performance overall and on both easy and difficult subsets predicted performance on their counterparts at retention. Within training conditions, the only significant predictor was the overall score within the difficult-first group.

To examine the ability of final acquisition level to predict retention level, correlations on the test scores (overall scores, easy subset scores, and difficult subset scores) were examined both across and within initial training conditions. Across groups, the overall score and both subset scores on the Session 3 posttest predicted their counterparts on the retention test (overall score r-square = .68, p = .002; easy subset r-square = .65, p = .003; difficult subset r-square = .54, p = .010). Within each group the number of subjects is small. Within the difficult-first group, with three subjects at retention, the squared correlations were greater than .95 for all three scores, but the only correlation that reached significance was on the overall score, r-square = .999, p = .015. Within the easy-first group, with four subjects, the only correlation that approached significance was overall score, r-square = .87, p = .065, with the squared correlations for both of the subset scores at or below .15. Finally, within the all-first group, with four subjects, only the easy subset score approached significance, r-square = .88, p = .065 (for the difficult subset, r-square = .65, p = .195; for the overall score r-square = .75, p = .135).

Summary

None of the initial training conditions lent a lasting advantage during acquisition, although initial training did lend each part-set group an immediate temporary advantage on its practiced set. There was a trend suggesting a Retention pretest disadvantage for the difficult-first group, bu this disadvantage was not significant in a planned analysis of that test alone. Accuracy at retention reflected significant forgetting. Performance both during acquisition and at retention confirmed the a prior classification of subset difficulty.Across initial training conditions, final acquisition performance overall and on both easy and difficult subsets predicted performance on their counterparts at retention. Within training conditions, the only significant predictor was the overall score within the difficult-first group.

CHAPTER III

EXPERIMENT 2: THE EFFECT OF INITIAL TRAINING SUBSET DIFFICULTY ON ACCURACY AND REACTION TIME FOR MORSE CODE RECEPTION

Because Experiment 1 included retrospective reports, the reaction times for that experiment were not analyzed. It was reaction time that showed an advantage for the difficult-first group in the Pellegrino et al. (1991) study, so Experiment 2 was designed to examine the effects of initial training on reaction time in the Morse code reception task. This study included only the two part-set training conditions, easy-first and difficult-first, and the number of subjects in each condition was increased. Additional analyses were included in the predictive validity section to examine the predictive ability of beginning performance (Session 1 pretest).

Method

The design is a 2 (initial training condition) x 4 (session) x 2 (pretest/posttest) x 2 (code-letter difficulty) mixed factorial; the first factor was varied between subjects, and the other factors were varied within subjects. This is identical to Experiment 1's design but with only the two part-set initial training conditions and only one test order (test order did not have a significant main effect or interaction with initial training condition in Experiment 1). In this experiment, 13 subjects from the same pool as used in Experiment 1 were assigned to each initial training condition. The procedure was identical to that of the two initial part training conditions in Experiment 1.

Results and Discussion

All analyses of accuracy performance were completed using individuals' proportions of correct responses on the tests; the analyses were then redone using arcsines of the square root of the proportions correct for comparison. All significant/nonsignificant designations (alpha = .05) were the same for the two methods except where noted. All accuracy means provided are untransformed proportions correct. Analyses of reaction time were completed using individuals' mean logarithm reaction times on correct items on the tests. All reaction time means provided are antilog; that is, they are ten to the power of the logarithmic mean.

Performance during acquisition. Neither initial training condition lent a lasting advantage in terms of either accuracy or reaction time during acquisition, although initial training did lend each group an immediate temporary advantage on its practiced set. Both reaction time and accuracy performance confirmed the a prior classification of subset difficulty.

The pattern of accuracy performance over training and retention for subjects in the different initial training conditions is illustrated in Figure 4. A four-way mixed analysis of variance was conducted on proportion of correct responses on the tests during acquisition, the first three sessions (see Table 4). The main effect of initial training condition was not significant, $F(1, 24) = 2.0$, $p = .164$, $MSe = .020$. Improvement in performance due to practice was evident, with the main effects of session and of pretest/posttest both significant. Accuracy performance on the difficult code-letter pairs was significantly poorer than on the easy pairs, supporting the a priori classification of the stimulus subsets.

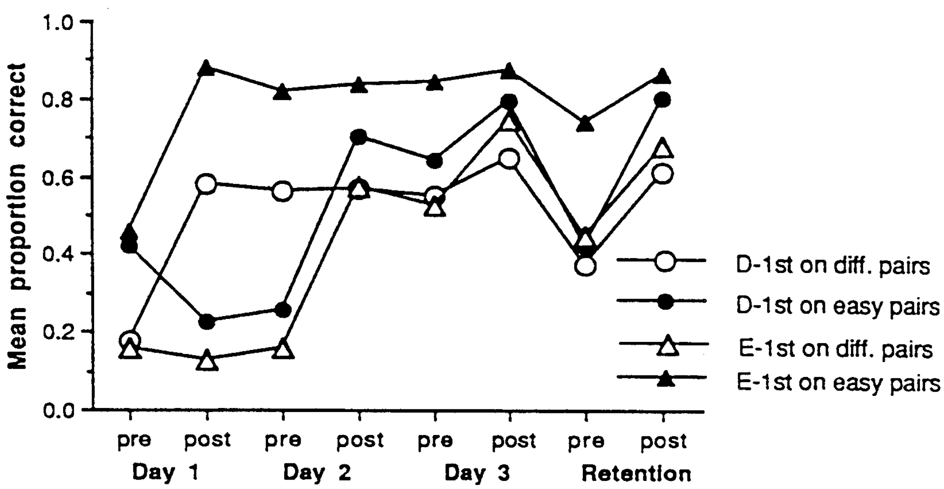


Figure 4. Experiment 2 mean accuracy over acquisition and retention sessions on the easy and difficult code-letter pairs

Table 4

Experiment 2 Proportion Correct ANOVA Results for All Effects with $p < .05$ on Acquisition (Arcsine p Values in Parentheses Where They Differ on Significance/Nonsignificance)

Variable	df	F	p	MSe
Session	2, 48	153.1	<.001	.018
Pretest/posttest	1, 24	128.9	<.001	.017
Letter difficulty	1, 24	123.3	<.001	.025
Initial training condition x letter difficulty	1, 24	132.6	<.001	.025
Session x pretest/posttest	2, 48	5.2	.009 (.052)	.013
Initial training x session x ltr difficulty	2, 48	38.8	<.001	.012
Initial training x session x pre/post x ltr difficulty	2, 48	95.6	<.001	.016

That the initial training did temporarily affect performance is evidenced by the significant interactions of letter difficulty and training condition; of session and pretest/posttest; of training condition, session and letter difficulty; and of training condition, session, pretest/posttest and letter difficulty. Directly after their initial partial training both groups showed better performance for the subset they had practiced than for the subset they had not, with the interaction between training condition and letter difficulty significant in post hoc analyses of both the Session 1 posttest and the Session 2 pretest, $F(1, 24) = 221.9$, $p < .001$, $MSe = .018$, and $F(1, 24) = 129.2$, $p < .001$, $MSe = .023$, respectively (with a Scheffe adjustment using m = 4)

A planned analysis on accuracy at the Session 3 posttest showed that initial training condition did not significantly affect final level of acquisition, $F(1, 24) = 1.8$, $p = .187$, $MSe = .051$. Final acquisition level also did not show a significant influence of training condition on the effect of letter difficulty as it had immediately after initial training, $F(1, 24) < 1$.

The pattern of reaction time performance over training and retention is illustrated in Figure 5. A four-way mixed ANOVA was conducted on mean log reaction times for the first three sessions of the experiment (see Table 5). Four subjects had no correct answers for their unpracticed subset on the Session 1 posttest and/or the Session 2 pretest, and one subject had no correct answers on the difficult subset of the Session 1 pretest; thus, correct mean log reaction times could not be computed for these cells. In these cases each missing value was replaced by the nearest extant value for the subject on that subset for the ANOVA.

Table 5

Experiment 2 Log Reaction Time ANOVA Results for All Effects with $p < .05$ on Acquisition

Variable	df	F	p	MSe
Session	2, 48	4.8	.012	.017
Letter difficulty	1, 24	113.9	< .001	.028
Initial training condition x letter difficulty	1, 24	15.9	< .001	.028
Pretest/posttest x letter difficulty	1, 24	5.0	.034	.002
Initial training x session x ltr difficulty	2, 48	5.4	.008	.008
Initial training x session x pre/post x ltr difficulty	2, 48	19.8	< .001	.005

Results largely parallel the findings for accuracy. As with accuracy, the main effect of initial training condition was not significant, $F(1, 24) < 1$. Speed-up in performance over the three sessions was statistically significant, although speed-up between the pretests and posttests was not. Responses on the difficult code-letter pairs were significantly slower than on the easy pairs, again supporting the a priori classification of the stimulus subsets.

Initial training condition temporarily affected reaction time as evidenced by the significant interactions of initial training condition and letter difficulty; of letter difficulty and pretest/posttest; of initial training condition, session, and letter difficulty; and finally of all four variables. After their initial subset training both groups showed greater improvement in

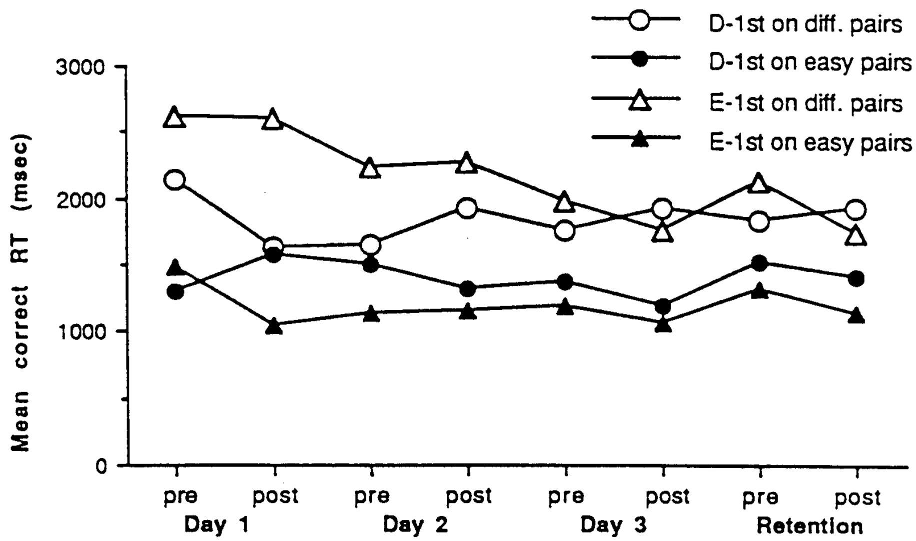


Figure 5. Experiment 2 mean reaction times over acquisition and retention sessions on easy and difficult code-letter pairs

reaction time for the subset they had practiced than for the subset they had not, with this interaction between training condition and letter difficulty significant in post hoc analyses of both the Session 1 posttest and the Session 2 pretest, $F(1, 24) = 23.3$, $p < .001$, MSe = .013, and $F(1, 24) = 10.6$, $p < .050$, MSe = .016, respectively (with a Scheffe adjustment using m = 4).

A planned analysis of reaction times on the Session 3 posttest showed that initial training condition did not significantly affect final level of acquisition, $F(1,24) = 2.0$, $p = .177$, MSe = .026. Final acquisition level also did not show the significant influence of training condition on the effect of letter difficulty that it showed immediately after initial training, $F(1, 24) < 1$.

Performance at retention. On the Retention pretest the easy-first group had a large advantage over the difficult-first group on the easy subset of code-letter pairs. The difficult group's small advantage on difficult items was not significant. Each group was significantly faster on the subset it initially practiced than was the other group. There was significant forgetting, reflected in lower accuracy, over the retention interval. Finally, both accuracy and reaction time performance supported the a prior classification of subset difficulty.

To examine the role of initial training subset on retention accuracy, a four-way mixed ANOVA was conducted on proportion correct for only the last two sessions, Session 3 and the Retention Session (see Table 6). The main effect of initial training condition was not significant, $F(1, 24) = 3.5$, $p = .071$, MSe = . 017. As would be expected, test scores at retention were lower than those at Session 3, and scores were higher on the posttests than on the pretests. As in acquisition, accuracy was lower for the difficult subset than for the easy subset.

Table 6

Experiment 2 Proportion Correct ANOVA Results for All Effects with $p < .05$ on Retention (Arcsine p Values in Parentheses Where They Differ on Significance/Nonsignificance)

Variable	df	F	p	MSe
Session	1, 24	16.6	< .001	.023
Pretest/posttest	1, 24	124.9	< .001	.014
Letter difficulty	1, 24	58.7	< .001	.027
Initial training condition x letter difficulty	1, 24	5.8	.023	.027
Session x pretest/posttest	1, 24	19.3	< .001	.009
Initial training x session x pretest/posttest	1, 24	5.4	.027 (.104)	.009
Initial training x pretest/posttest x ltr difficulty	1, 24	36.9	< .001	.006

The significant interaction between session and pretest/posttest reflects the greater dip in scores for the Retention pretest, which was the first test after 4 weeks without practice, than for the Session 3 pretest, which was in the middle of acquisition.

The interaction between training condition and letter difficulty was significant, as were the three-way interactions of training condition, pretest/posttest, and letter difficulty; and of training condition, session and pretest/posttest. Separate analyses of each of the tests revealed that the interaction of training condition and letter difficulty was significant only on the pretests ($F(1, 24) = 17.7$, $p < .005$, $MSe = .010$, on a post hoc analysis of the Session 3 pretest with Scheffe adjustment using m = 4 ; $F(1, 24) = 9.8$, $p = .005$, $MSe = .019$ on a planned analysis of the Retention pretest). On both pretests the easy-first group showed better performance on the easy subset than did the difficult-first group, but the two groups were nearly the same on difficult-first items.

A planned analysis of performance on only the Retention pretest yielded significant effects of training condition and letter difficulty as well as of their interaction. Overall, the easy-first group scored more highly, $F(1, 24) = 9.2$, $p = .006$, $MSe = .053$; and accuracy remained higher on the easy subset than on the difficult subset, $F(1, 24) = 19.9$, $p < .001$, $MSe = .019$. As the significant interaction between letter difficulty and training condition suggests, the primary advantage was for the easy-first group over the difficult-first group on the easy subset of code-letter pairs, $F(1, 24) = 9.8$, $p = .005$.

The effect of initial training subset on reaction time at retention was examined with a four-way mixed ANOVA on the last two sessions, Session 3 and the Retention Session (see Table 7). The main effect of initial training condition was not significant, $F(1, 24) < 1$. Reaction times during the Retention Session were not significantly different from those during Session 3, but times were shorter on the posttests than on the pretests. As in acquisition, response time was significantly slower for the difficult subset than for the easy subset.

Table 7

Experiment 2 Log Reaction Time ANOVA Results for All Effects with $p < .05$ for Retention

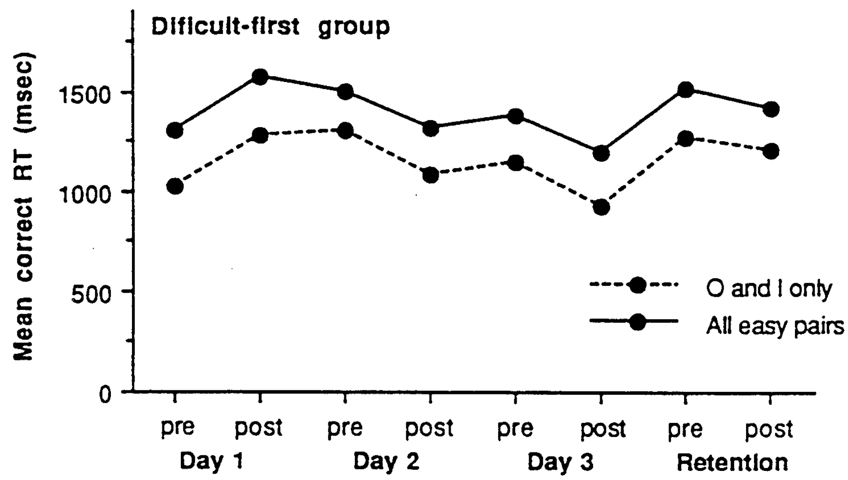
Variable	df	F	p	MSe
Pretest/posttest	1, 24	5.6	.025	.012
Letter difficulty	1, 24	105.5	< .001	.014
Initial training condition x letter difficulty	1, 24	5.0	.034	.014
Session x letter difficulty	1, 24	6.3	.018	.003
Initial training x pretest/posttest x ltr difficulty	1, 24	6.0	.021	.004

There is a significant interaction between session and letter difficulty; for both sessions responses are slower for the difficult subset than for the easy subset, but this is especially the case for the Retention session. The interaction between training condition and letter difficulty was significant, as was the three-way interaction of training condition, pretest/posttest, and letter difficulty. On difficult code-letter pairs the difficult-first group responded more quickly than the easy-first group on the pretests, but the easy-first group responded more quickly on the posttests. On the pretests each group therefore showed better performance on their initially trained subset than did the other group, but on the posttests the easy-first group was faster than the difficult-first group on both subsets. This interaction of training condition and letter difficulty was significant in a planned analysis of the Retention Session pretest ($F(1, 24) = 6.3$, $p = .021$, $MSe = .010$) but not in a post hoc analysis of the Session 3 pretest ($F(1, 24) = 7.0$, $p > .090$, $MSe = .005$, with a Scheffe adjustment using m = 4).

The planned analysis of performance on the Retention pretest also yielded a significant effect for letter difficulty. Overall, reactions remained faster on the easy subset than on the difficult subset, $F(1, 24) = 26.8$, $p < .001$, $MSe = .010$.

Representativeness of Correct Reaction Time Analyses. Because the above reaction time analyses were completed on only correct responses, there was a possibility that the analyzed pattern of reaction times was an artifact; that is, it was possible that later mean reaction times were artificially slower because they included reaction times on newly-correct, and thus more slowly identified, items. However, analysis did not support that possibility.

Because I and O were the most-frequently correct letters across subjects, a four-way mixed ANOVA was completed on the whole easy subset versus the easy letters I and O only (see Table 8 and Figure 6). Two subjects had no correct answers for the easy subset (and necessarily for the letters I and O) on the Session 1 posttest and/or the Session 2 pretest, and two more subjects had no correct answers on those tests for the letters I and O; thus, correct mean log reaction times could not be computed for these cells. In these cases each missing value was replaced by the nearest extant value for the subject on that set for the ANOVA. Although subjects were significantly faster on the letters I and O than on the whole easy subset, there was neither an interaction with session ($F(3, 72) < 1$) nor an interaction with pretest/posttest ($F(1, 24) = 3.0$, $MSe = .003$, $p = .092$) nor with both ($F(3, 72) = 1.6$, $MSe = .004$, $p = .184$).



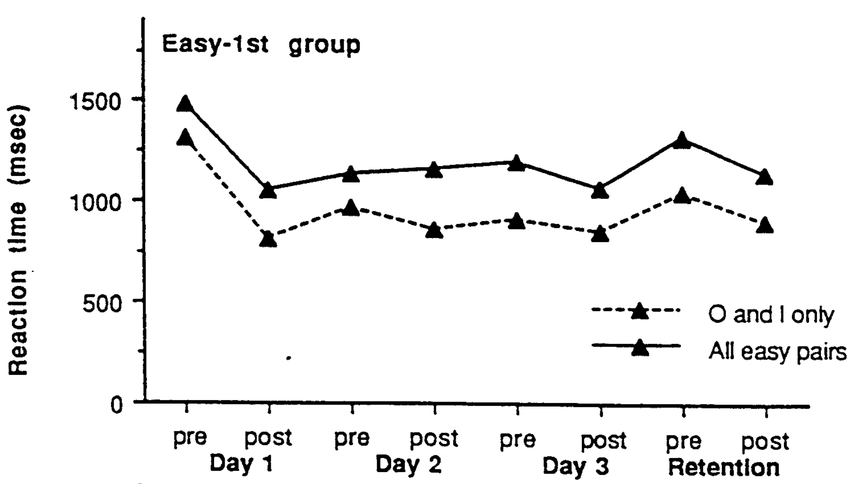


Figure 6. Experiment 2 change in mean RT over acquisition and retention for O and I versus for all easy pairs

Table 8

Experiment 2 Easy/ I, O Log Reaction Time ANOVA Results for All Effects with $p < .05$

Variable	df	F	p	MSe
Easy subset/ I, O only	1, 24	85.0	< .001	.009
Session	3, 72	4.0	.012	.015
Pretest/posttest	1, 24	8.7	.007	.028
Initial training x session x pre/post	3, 72	12.2	< .001	.004

Predictive validity. For both accuracy and reaction time, final acquisition performance overall and on both easy and difficult subsets predicted performance on their counterparts at retention when examined across training conditions or within the easy-first initial training group. However, within the difficult-first group only the performance on the difficult subset predicted its counterpart at retention. Retention and final acquisition level predictions based on beginning performance were valid for some measures but not within the easy-first group.

To examine the ability of final acquisition level to predict retention level, correlations on the test scores were examined both across and within initial training conditions. Across groups, the overall score and both subset scores on the Session 3 posttest predicted their counterparts on the retention test (overall score r-square = .38, $p < .001$; easy subset r-square = .26, $p = .008$; difficult subset r-square = .43, $p < .001$). The same pattern of predictive ability was found within the easy-first training group (overall score r-square = .66, $p < .001$; easy subset r-square =.63, $p = .001$; difficult subset r-square = .54, $p = .004$). Within the difficult-first group, however, only the performance on the difficult subset predicted its counterpart at retention, r-square = .43, $p = .014$.

Correlations between Session 3 posttest mean reaction times and Retention pretest mean reaction times paralleled the correlations for accuracy performance. Across initial training conditions the mean reaction times overall and on both subsets predicted their counterparts on the retention test (overall r-square = .37, $p < .001$; easy subset r-square = .42, $p < .001$; difficult subset r-square = .28, $p = .006$). The same pattern of predictive ability was found within the easy-first group (overall r-square = .72, $p < .001$; easy subset r-square = .60, $p = .002$; difficult subset r-square = .66, $p < .001$). Within the difficult-first group, only the mean reaction time on the difficult subset predicted its counterpart at retention, r-square = .45, $p = .013$.

To examine the ability of the Session 1 pretest to predict final acquisition accuracy (Session 3 posttest) and retention accuracy (Retention pretest), correlations on the test scores were examined both across the initial training conditions and within them. Across training conditions performance on the easy subset of the Session 1 pretest predicted final acquisition performance on that subset, r-square = .22, $p = .015$, but did not predict retention level, r-square = .09, $p = .132$. Session 1 pretest performance on the difficult subset predicted only retention level, r-square = .19, $p = .028$.

Within the difficult-first training group, performance on the easy subset at the Session 1 pretest predicted final acquisition level on that subset, r-square = .45, $p = .012$. Beginning overall test score across letter difficulty was the only significant predictor of retention level, predicting overall retention score with an r-square = .35, $p = .034$.

Within the easy-first training group no score, not overall score nor easy subset score nor difficult subset score, predicted final acquisition level. Beginning performance on the difficult subset did marginally predict performance on that subset at retention, however, r-square = .29, p = .057 (arcsines $p = .048$).

Reaction time correlations for the Session 1 pretest were also examined. Across initial training conditions, beginning reaction time predicted final acquisition level only for the difficult subset, r-square = .18, $p = .034$. However, beginning performance on all three measures, overall and both subsets, predicted their counterparts at retention (overall score r-square = .29, p = .006; easy subset r-square = .38, p = .001; difficult subset r-square = .21, p = .021).

Within the difficult-first group, reaction time on the full set of stimuli on the Session 1 pretest predicted retention level overall (r-square = .40, p = .020). Beginning mean reaction time on the difficult subset also predicted both final acquisition and retention performance on that subset (r-square = .33, p = .039; r-square = .32, p = .045; respectively). Within the easy-first group, none of the beginning reaction times predicted either final acquisition level or retention level.

Summary

Neither initial training condition lent a lasting advantage in terms of either accuracy or reaction time during acquisition, although initial training did lend each group an immediate temporary advantage on its practiced set. On the Retention pretest the easy-first group had a large advantage over the difficult-first group on the easy subset of code-letter pairs. The difficult group's small advantage on difficult items was not significant. Each group was significantly faster on the subset it initially practiced than was the other group. There was significant forgetting, reflected in lower accuracy, over the retention interval. Finally, both accuracy and reaction time performance supported the a prior classification of subset difficulty both during acquisition and at retention.

For both accuracy and reaction time, final acquisition performance overall and on both easy and difficult subsets predicted performance on their counterparts at retention when examined across training conditions or within the easy-first initial training group. However, within the difficult-first group only the performance on the difficult subset predicted its counterpart at retention. Retention and final acquisition level predictions based on beginning performance were valid for some measures but not within the easy-first group.

CHAPTER IV

EXPERIMENT 3: SUBTASK TRAINING OF MORSE CODE RECEPTION

The goal of Experiment 3 was to localize the sources of difficulty for all the Morse code stimuli. For example, it is possible that the major cause of difficulty on the difficult subset of items is in perceiving the auditory code properly; at a lower level of analysis it is possible that errors in perceiving the code are always manifested by responding with an incorrect letter whose code is of the same length has the presented code. To localize the sources of difficulty, the Morse code reception task was divided into parts in terms of subtasks. Based on pilot protocol work, the task of Morse code reception appears to consist of two natural subtasks, segmenting the signal into its component elements (dots and dashes), then accessing the arbitrary letter that corresponds to that pattern of elements. This experiment addresses whether these two stages contribute differently to the errors made during Morse code reception.

Research on Morse confusion errors, errors of hearing one code but writing the letter of another code, reaches back to World War II. The earliest studies (Keller & Taubman, 1943; Plotkin, 1943; Spragg, 1943) consist mainly of tabulating confusions errors, rank ordering the difficulty of the Morse codes and determining groups of highly confusable letters. Gibson (1969) examining the Morse confusion data of Keller & Taubman noted that errors based on similarity of letter sounds (for example, A and K) were rare. In 1957, Rothkopf hypothesized, based on a same-different discrimination task, that similarity predicts Morse code reception performance, but did not further define similarity.

A large study by Seashore and Kurtz (1944) examined pairs of most-likely-to-be-confused codes and concluded that there were five types of reception error. The most common error was shortening the signal, either substituting a dot for a dash in the signal or dropping a code element (a dot or a dash) usually at the end of the code. Other error types were lengthening the signal, completely substituting dots and dashes, altering internal elements within the signal, and a miscellaneous category. Highland and Fleishman (1958) carried out a factor analysis on confusion errors produced by Morse operator trainees, finding five factors. The most common was "dot estimation", adding or deleting a dot in the code; the next most common was end-element substitution, flipping the last element in the code from a dot to a dash or vice versa; three remaining factors were dash estimation, altering an internal element within the signal, and a final factor accounting solely for the confusions between the letter V and the number 4.

Finally, Shepard in 1963 applied his multivariate analysis to the data of Rothkopf, of Keller and Taubman, of Plotkin, and of Seashore and Kurtz. Shepard divided these studies into three groups. First, Rothkopf's discrimination data suggested that signals were confused on the basis of two factors, the number of elements in the codes and the predominance of dashes or dots in the codes. Second,the work both of Keller and Taubman and of Plotkin gave measures of the confusion errors in novice Morse operators. Novice errors could be explained by two factors: again the number of elements in the codes and this time the heterogeneity of the codes (that is, whether the code was made up of purely one type of element or whether it was mixed). Finally, the Seashore and Kurtz data focused on advanced Morse receivers. For these operators, there was but one strong factor: dropping or adding a dot or dash in a series. For example, typing the letter A (. _) when hearing the code for U (. . _).

These studies all classified the errors a posteriori. Experiment 3 uses an a priori classification that is functional in nature, necessarily including all possible confusion errors for the set of 2- and 3-element codes. The errors are broadly classified into either same-length errors (those in which the code sent and the code for the typed character have the same number of elements) or different-length errors. The same-length errors are further classified as those in which the sent and response codes differ by one element ("1-different" errors), those that differ by two elements ("2-different" errors), and those that are opposites ("opposite" errors).

To examine the effects of the different subtasks on errors, three groups of subjects received training on different tasks. The first group (the code-to-letter group) trained on the complete Morse code reception task of hearing codes and typing their corresponding letters. The second group (the code-to-dida group) trained on the first subtask, segmenting the codes into their elements, hearing codes then responding by typing keys labelled "di" (for a dot) and "da" (for a dash) for the elements they heard. The third group (the dida-to-letter group) trained on the second subtask, translating the pattern of elements into a letter, viewing "dida" patterns (e.g. "didadi" for the letter R) on the CRT then typing their corresponding letters. These dida patterns were used in place of the standard periods and hyphens in order to reduce possible visual pattern processing, because visual pattern processing would not be part of the letter-accessing subcomponent of auditorily presented Morse code reception. Subjects completed a pretest, two sessions of training, and a posttest; two weeks after the posttest the subjects returned for a retention test. All training and tests covered only the task (code-to-letter, code-to-dida, dida-to-letter) assigned.

Method

Design. The design used was a 3 (task) x 3 (test time-- pretest, posttest, and retention test) x 2 (number of elements in the code) mixed factorial design. The first factor, task, was varied between subjects. The remaining factors -- test time and number of elements in the code -- were varied within subjects.

Subjects. The subjects were students in introductory psychology classes at the University of Colorado, Boulder, who participated in order to fulfill a class requirement. All were native

English speakers who did not have previous knowledge of Morse code. There were seven subjects per task type. Assignments were made on the basis of a fixed rotation, according to the time of arrival for the first session.

Materials. For the code-to-letter and code-to-dida groups, the codes were presented aurally by an IBM PC; for the dida-to-letter group, the dida patterns were displayed on the CRT. For the code-to-letter and dida-to-letter groups, subjects typed their responses on a normal PC keyboard; the keys for the twelve letters in the stimulus set were marked with black tape to distinguish them from the letters not to be used. Subjects in the code-to-dida group typed their responses using two keys on the keyboard that were marked "di" and "da". For all subjects, the Return key was also marked with black tape.

Subjects were trained on the full set of two- and three-element codes (see Figure 1 of Experiment 1). The three tests were identical, consisting of four blocks of the 12 randomly arranged codes. Training was also presented in randomly arranged blocks of 12.

Procedure. The initial session of training began with orientation tasks. The subjects first were given four blocks of training simply typing the 12 responses that they would later use in their tasks followed by the return key; for the code-to-letter and dida-to-letter groups, this was practice typing the 12 letters, and for the code-to-dida group this practice was on typing the 12 dida patterns. For all three tasks, after the subject hit the return key, the computer gave accuracy feedback ("Correct response!" or "Wrong response!") and , if the subject was correct, the initial reaction time in seconds. After a two-block typing test (without feedback) on the practiced responses, the subject began learning the assigned task. First, the computer cycled through the full set of twelve pairs four times, playing the code while showing its corresponding letter on the screen, or playing the code while showing its dida pattern on the screen, or showing a dida pattern and its letter on the screen, depending on the subject's task. After this introduction to the pairs, the subject was presented with the pretest for the assigned task.

After the pretest the subjects received 120 trials of practice on their task. Two days later the subjects returned for another 180 trials of training followed by the posttest. Fourteen days later the subjects returned for the retention test.

The procedure for a training trial was similar for all three tasks. For the code-to-letter subjects the computer displayed the phrase "Type the letter for this code," played a Morse code, then prompted the subject ("Now") to type the corresponding letter followed by the return key. The code-to-dida subjects saw the phrase "Type the di-da pattern for this code," heard the code, then were prompted ("Now") to type a series of "di"s and "da"s to make the pattern of elements that they had heard, followed by a return. The dida-to-letter subjects saw the phrase "Type the letter for this di-da pattern," saw a pattern of dis and das on the screen, then were prompted ("Now") to type the letter corresponding to that pattern, followed by a return.

In all three tasks, after the subject hit the return key, the computer gave feedback. If the subject was correct, the computer responded with "CORRECT" and the initial reaction time in seconds; if the subject was incorrect, the computer responded with "INCORRECT" and displayed the correct response (in the phrase "That was the di-da pattern for ..." or "That was the code for ..."). Finally, regardless of the subject's accuracy, the computer reviewed the trial by again playing the stimulus code or displaying the stimulus dida pattern, then displaying the correct response.

The test trial procedure was identical to the training trial procedure except that there was no feedback and no review.

Results and Discussion

Because the results of accuracy analyses in Experiment 2 were similar whether completed on proportions correct or on arcsines of the proportions, accuracy analyses for this experiment were completed on proportions correct only. All accuracy means provided are untransformed proportions correct.Analyses of reaction time were completed using individuals' mean logarithm reaction times on correct items on the tests. All reaction time means provided are antilog; that is, they are ten to the power of the logarithmic mean.

Effect of task and number of elements on performance. Task did not affect accuracy although it affected reaction times with the code-to-dida consistently faster than the other groups. Overall, reaction times and accuracy both suggested that 2-element codes are easier than 3-element codes and that performance improved with practice then worsened over retention. These effects were influenced by task; most noticeably, the code-to-dida group showed consistency across tests both in terms of accuracy and of reaction times with no loss over retention.

The pattern of accuracy performance over training and retention for subjects in the three task groups is illustrated in Figures 7 and 8. A three-way mixed analysis of variance was conducted on accuracy (see Table 9). The main effect of task was not significant, $F(2, 18) = 2.5$, $p = .107$, $MSe = .166$. Across tasks, improvement with practice and decline over retention were reflected in the significant main effect of test. Subjects exhibited lower accuracy on 3-element codes than on 2-element codes. These last two effects were influenced by task, however. First, the code-to-dida group showed slighter improvement over practice and no loss over retention, leading to a significant interaction of test and task (note that the code-to-dida group did show significant learning; a post hoc analysis revealed that the differences between posttest and pretest proportions were significantly different from zero, with a mean improvement of .12, $t(1, 6) = 3.2$, $p < .05$ with a Scheffe adjustment using $m = 3$). Second, the influence of test on the effect of elements was different for the three tasks. For the code-to-dida group, the advantage for 2-element codes remained relatively stable throughout the experiment. For the code-to-letter group, the 2-element advantage developed after practice; there was no advantage on the pretest, only on the posttest and the retention test. For the dida-to-letter group the advantage decreased at the

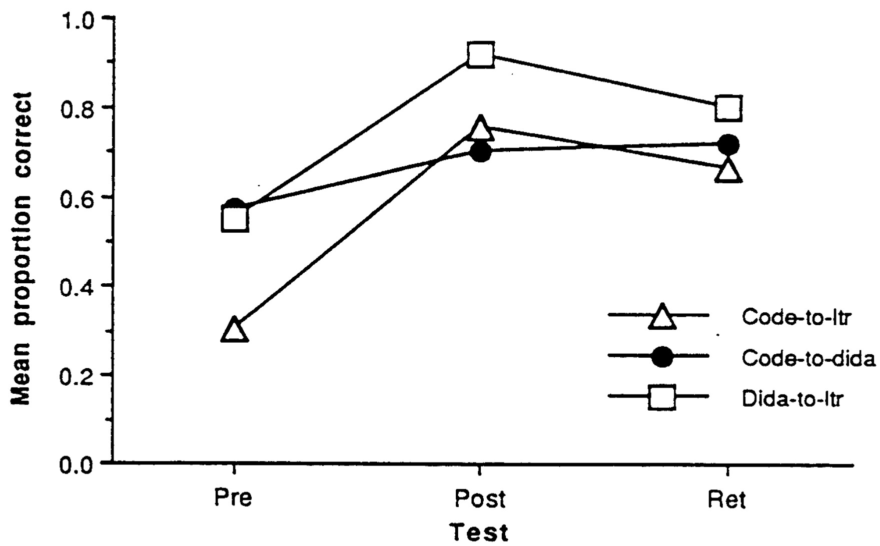


Figure 7. Experiment 3 mean accuracy over acquisition and retention sessions

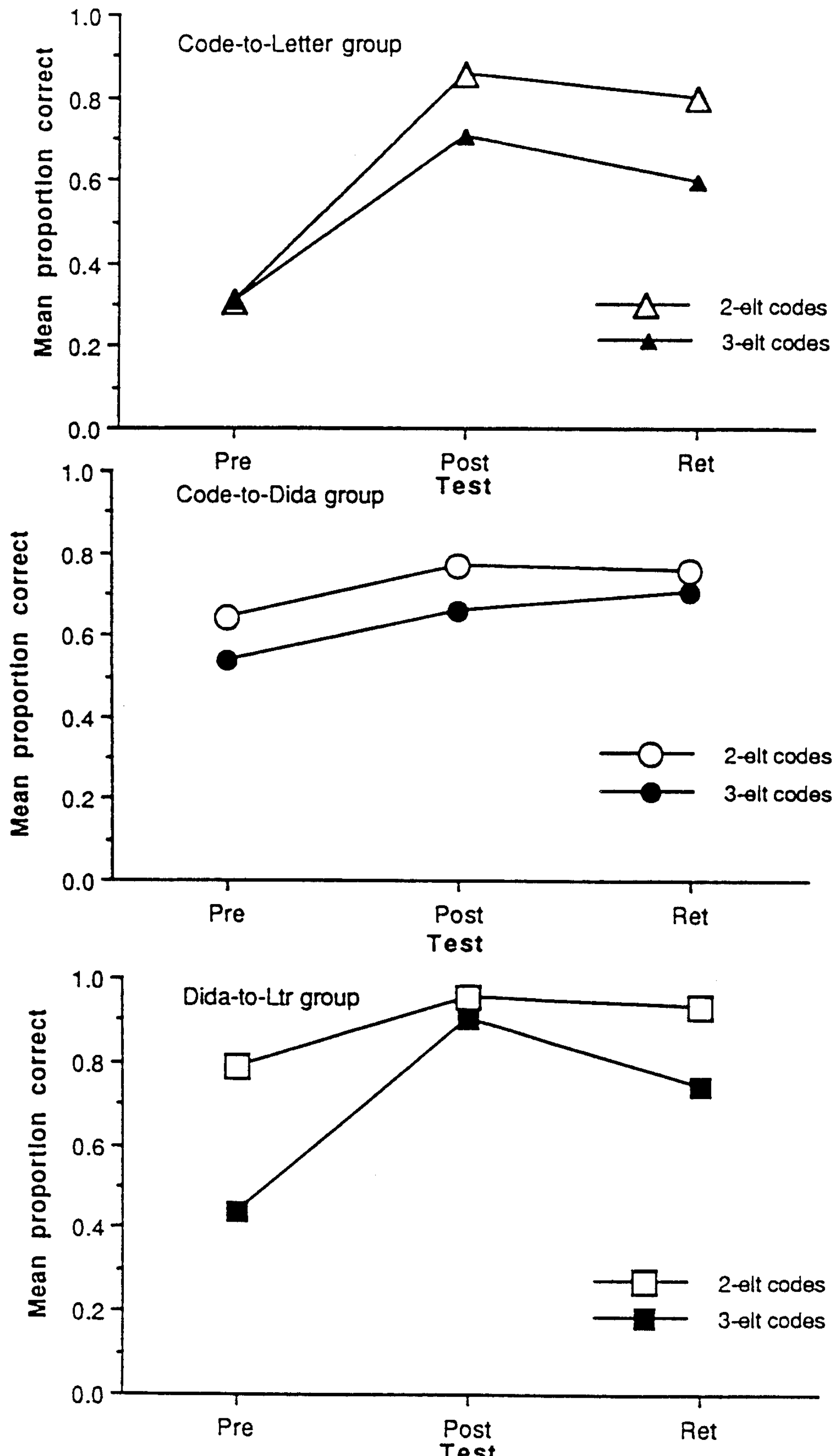


Figure 8. Experiment 3 mean accuracy over acquisition and retention sessions for 2- and 3-element codes

posttest then increased again over retention; practice decreased the advantage for 2-element codes.

Table 9

Experiment 3 Proportion Correct ANOVA Results for All Effects with $p < .05$

Variable	df	F	p	MSe
Test	2, 36	48.4	< .001	.023
Elements	1, 18	29.7	< .001	.019
Task x test	4, 36	5.2	.002	.023
Task x test x elements	4, 36	5.6	.002	.010

A planned analysis of the effect of task on accuracy at posttest showed that the effect of task was not significant, $F(2, 18) = 2.2$, $p = .139$, $MSe = .042$. At retention the effect of task was again not significant, $F(2, 18) = 1.0$.

The pattern of reaction time performance over training and retention for subjects in the three task groups is illustrated in Figures 9 and 10. A three-way mixed ANOVA was conducted on reaction times (see Table 10). One subject had no correct answers on the 2-element codes of the pretest and therefore no mean correct reaction time; for the ANOVA that missing value was replaced with the subject's posttest mean correct reaction time on 2-element codes. Across tasks and elements, reaction times improved (i. e., declined) after training then increased at retention, and this main effect of test was significant. The main effect of task was significant, with the code-to-dida group initiating responses fastest, followed by the code-to-letter group and then the dida-to-letter group. This effect was significantly influenced, however, by test; the dida-to-letter group was slower than the code-to-letter group only on the pretest (the code-to-dida group remained fastest throughout the experiment). Responses were significantly faster on 2-element codes than on 3-element codes. This

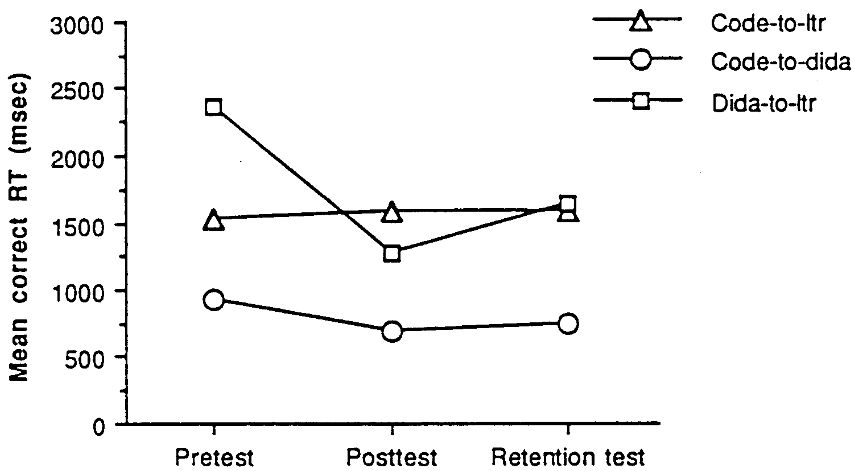


Figure 9. Experiment 3 mean reaction times over acquisition and retention sessions

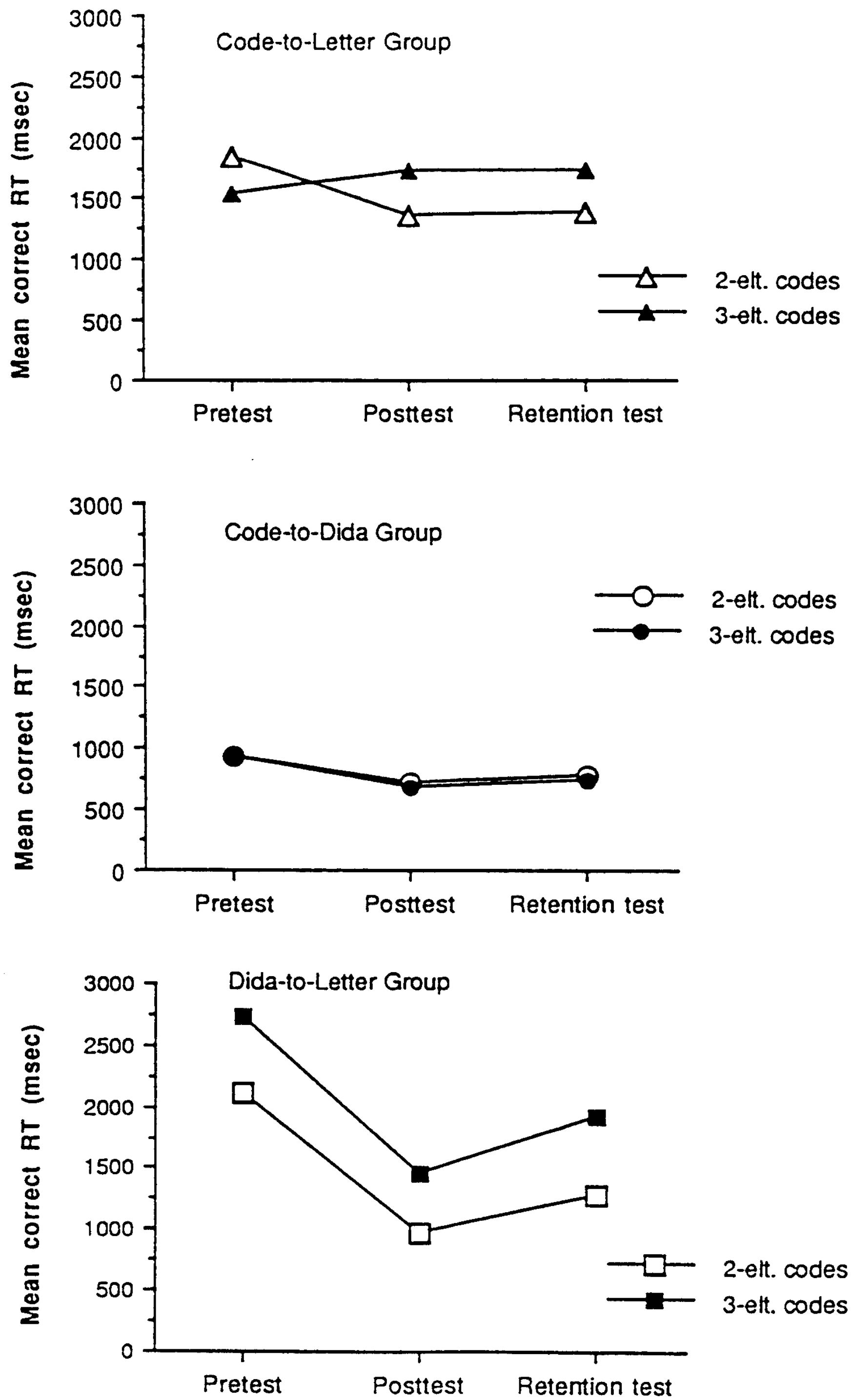


Figure 10. Experiment 3 mean reaction times over acquisition and retention sessions for 2- and 3-element codes

advantage for 2-element codes was not apparent, though, in the performance of the code-to-dida group, with the interaction between elements and task significant.

Table 10

Experiment 3 Log Reaction Time ANOVA Results for All Effects with $p < .05$

Variable	df	F	p	MSe
Task	2, 18	32.2	< .001	.042
Test	2, 36	14.2	< .001	.018
Elements	1, 18	16.9	< .001	.008
Task x test	4, 36	4.0	.009	.018
Task x elements	2, 18	10.5	.001	.008

A planned analysis of the effect of task on reaction time at posttest showed that the effect of task was significant, $F(2, 18) = 16.8$, $p < .001$, $MSe = .014$. On the retention test the effect of task was significant, $F(2, 18) = 24.2$, $p < .001$, $MSe = .011$. On both tests the code-to-dida group reacted over 500 msec faster than the other two groups. In post hoc analyses using a Scheffe adjustment (m = 3), the code-to-dida reaction times were significantly faster than the reaction times of the other two groups on all three tests (on the pretest $F(1, 18) = 17.6$, $p < .005$, $MSe = .025$; on the posttest $F(1, 18) = 31.4$, $p < .001$, $MSe = .014$; on the retention test $F(1, 18) = 48.3$, $p < .001$, $MSe = .011$). The code-to-letter group did not differ significantly from the dida-to-letter group on any of the tests.

Influence of Task on the Effect of Letter Difficulty. The effect of letter difficulty is not peculiar to only one of the subtasks. Subjects in all three tasks showed a performance advantage for the easy subset of stimuli, in terms of both accuracy and reaction time, although the reaction time advantage for the code-to-dida group was smaller than for the other groups.The pattern of accuracy performance over training and retention for subjects in the three task groups is illustrated in Figure 11. A three-way mixed analysis of variance was conducted on accuracy (see Table 11). The main effect of test as well as its interaction with task agree with the above analysis. Although there was a significant advantage for easy letters, an advantage that was greatest on the pretest, the effect of letter difficulty was not influenced by task ($F(2,18) < 1$) nor was there a significant three-way interaction ($F(4,36) < 1$).

Table 11

Experiment 3 Letter Difficulty Proportion Correct ANOVA Results for All Effects with $p < .05$

Variable	df	F	p	MSe
Test	2, 36	53.2	< .001	.022
Task x test	4, 36	4.9	.003	.022
Letter difficulty	1, 18	65.8	< .001	.016
Test x letter difficulty	2, 36	5.7	.007	.014

The pattern of reaction time performance over training and retention for subjects in the three task groups is illustrated in Figure 12. A three-way mixed ANOVA was conducted on reaction times (see Table 12). The main effects of task and test agree with the above analysis. As with accuracy, performance on easy letters was significantly better than on difficult letters. However, this effect was influenced by task, with the code-to-dida group showing the smallest advantage for the easy subset. A post hoc two-way repeated measures ANOVA (using a Scheffe adjustment with m = 5) on only the code-to-dida group showed that the easy subset advantage was still significant, $F(1,6) = 40.6$, $MSe = .005$, $p < .01$ (neither the effect of test, $F(2,12) = 2.3$, $MSe = .027$, $p > .1$, nor the interaction, $F(2,12) < 1$, was significant).

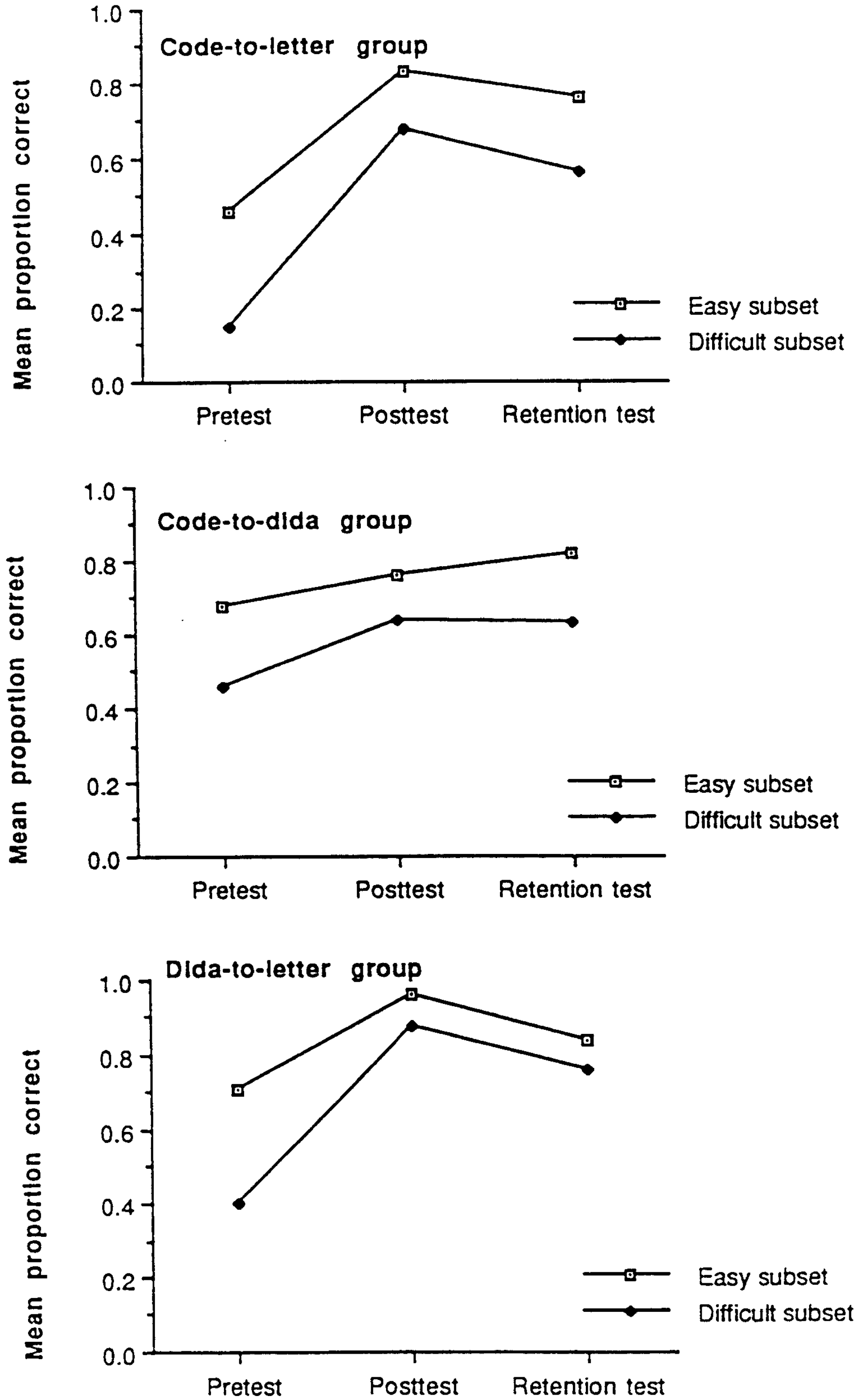


Figure 11. Experiment 3 letter difficulty mean accuracy over acquisition and retention sessions

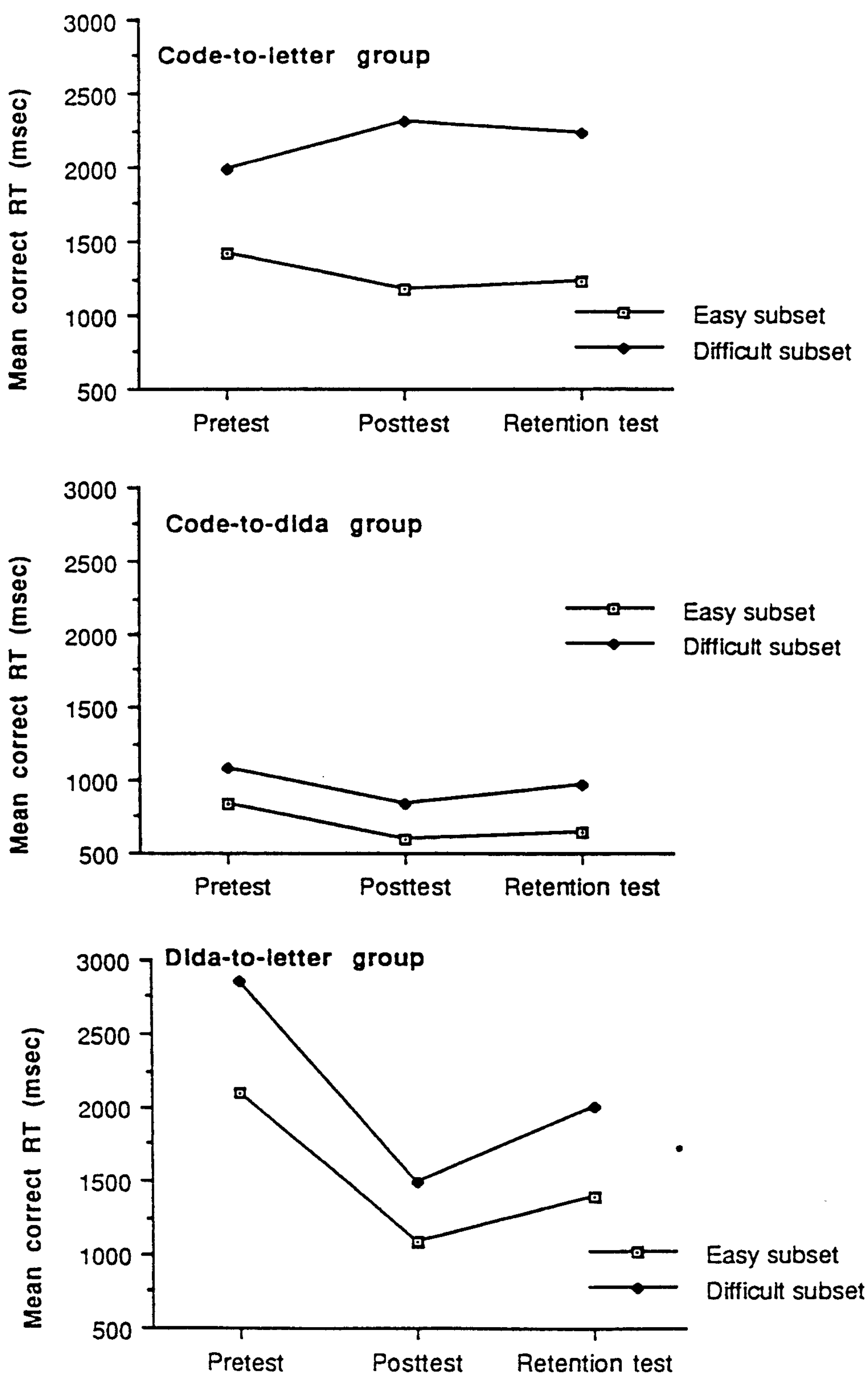


Figure 12. Experiment 3 letter difficulty mean reaction time over acquisition and retention sessions

Table 12

Experiment 3 Letter Difficulty Log Reaction Time ANOVA Results for All Effects with $p < .05$

Variable	df	F	p	MSe
Task	2, 18	23.1	< .001	.062
Test	2, 36	6.8	.003	.030
Letter difficulty	1, 18	126.0	< .001	.007
Task x letter difficulty	2, 18	4.1	.033	.007

Predicting complete task performance with subtask performance. Observed accuracy performance lent some support to the a priori decomposition of the complete task into the code-to-dida and dida-to-letter subtasks, although this decomposition is best understood as an approximation to the whole task. If the process for performing the complete Morse code reception task is composed of the two subtasks, code-to-dida and dida-to-letter, then accuracy on the two subtasks should predict accuracy on the complete task. That is, on the complete task some percentage of the codes would be segmented into their elements correctly; of those, some percentage would be associated with their letters correctly.To analyze whether accuracy on the subtasks predicted accuracy on the complete task, the mean accuracies for the code-to-dida and dida-to-letter groups were multiplied to yield a numerical prediction of accuracy for the complete task group. Figure 13 illustrates the relationship between the subtask-derived predictions and the observed means of the code-to-letter group. For each test, the subtask-derived prediction was compared to the performance of subjects in the code-to-letter group, and the differences were not significant on any of the tests (pretest $t(1, 6) < 1$; posttest $t(1, 6) = 1.4$, $p = .207$; retention test $t(1, 6) = 1.1$, $p = .334$).

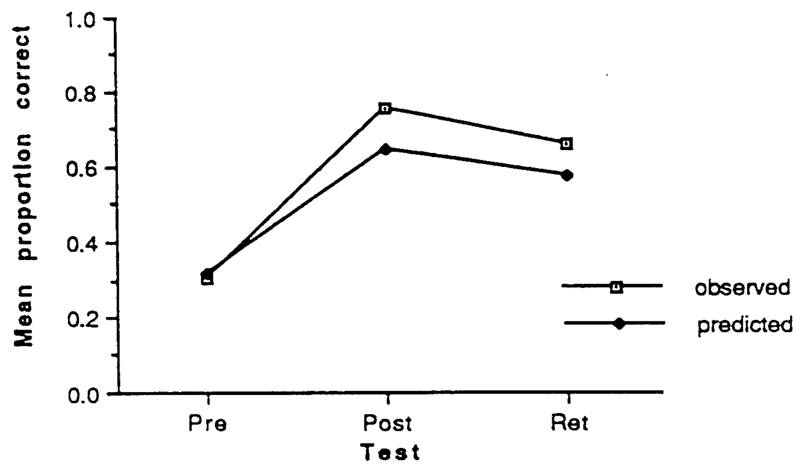


Figure 13. Experiment 3 observed and predicted mean accuracy for the code-to-letter task

If the Morse reception task is composed of the two subtasks, then also the sum of the error frequencies for the two subtasks should approximate the error frequencies for the whole task. Figures 14 and 15 display the ability of subtask performance to predict whole task performance. This information is only descriptive, with no statistical analyses. For these figures the numbers of errors of each type were summed across all subjects in each condition, then the sums for the code-to-dida group and the dida-to-letter group were added to become the predicted frequency; note that this method does not take into account the possibility that an error in segmenting the code could be corrected by an ensuing error in retrieving the corresponding letter. Across number of elements and tests the predicted frequencies were greater than those actually observed for 1-different errors, and on the 2-element codes of the pretest the predicted frequencies were less than those observed for different-length errors.

There are a number of reasons why subtask performance might not predict whole task performance. Among them is the possibility that subjects in the whole task may not be segmenting the codes they hear but may rather be perceiving the codes as rhythms; for example, the letter 'V' (which was not used in this experiment) is represented by . . . _ which is easily recognized without segmenting because it has the rhythm of the famous measures from Beethoven's Fifth Symphony. There is also the possibility that the presentation of the dida patterns in the dida-to-letter condition allowed mnemonic devices that are unavailable for the auditory codes; for example, the dida patterns for 'A' and for 'I' end in the letters 'a' and 'i,' respectively. More systemically, there may be interactions within the whole task that span the two separate subtasks and are therefore unobservable in those subtasks alone.

Predictive validity. Within the code-to-dida group, accuracy was consistently predictable throughout the experiment, and reaction time on the posttest predicted reaction time on the retention test. Posttest performance predicted retention test performance within the code-to-letter group for accuracy only and within the dida-to-letter group for reaction time only.

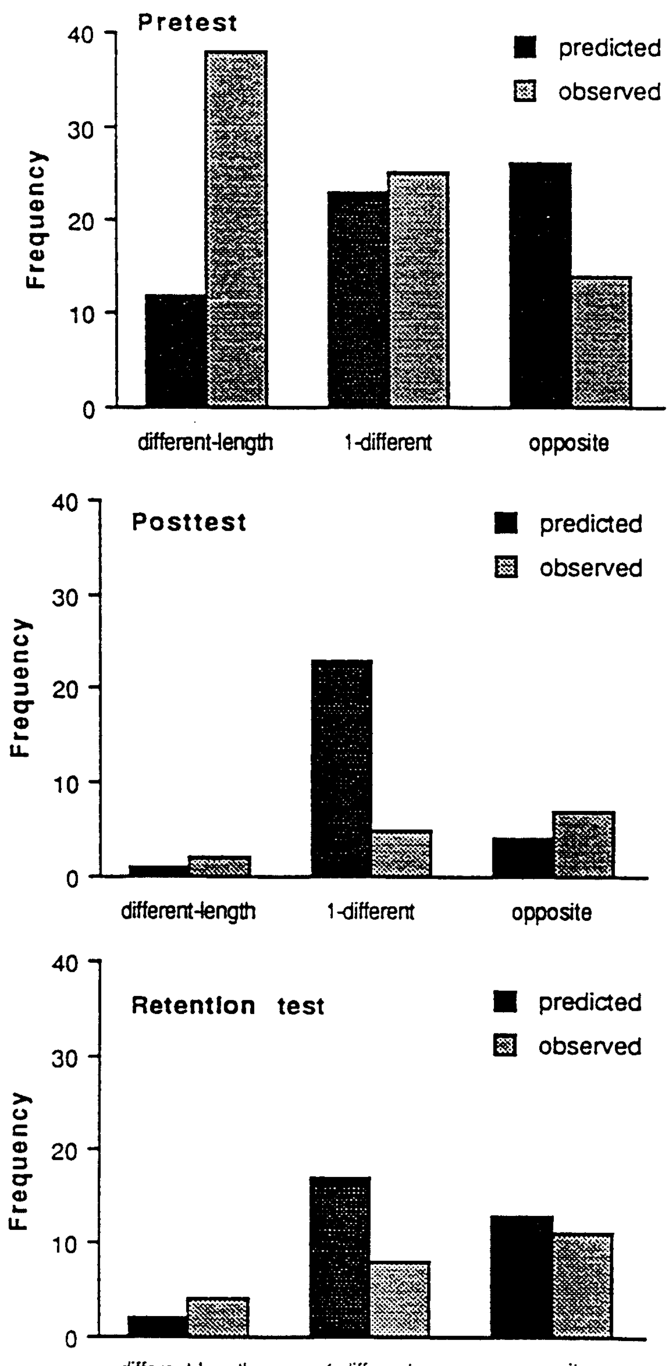


Figure 14. Experiment 3 observed and predicted error frequencies for the code-to-letter task on 2-element codes

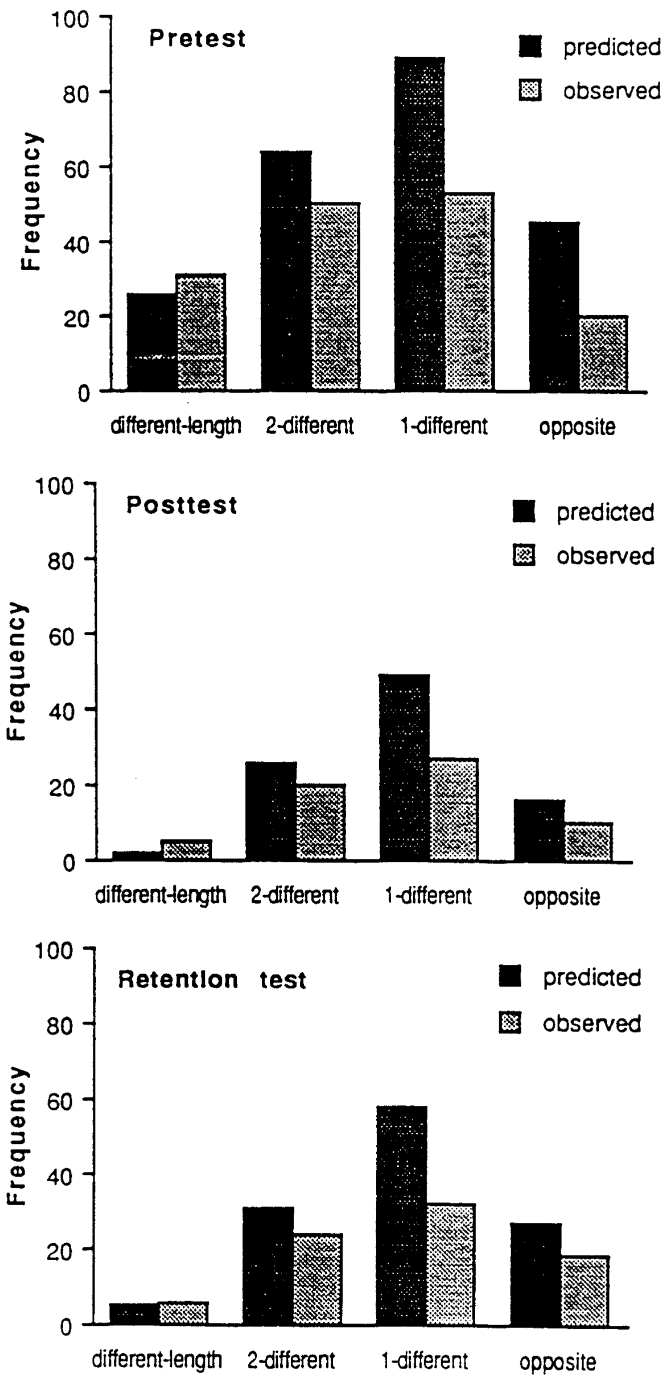


Figure 15. Experiment 3 observed and predicted error frequencies for the code-to-letter task on 3-element codes

To examine the ability of posttest accuracy to predict retention test accuracy , and to examine the ability of pretest accuracy to predict posttest and retention test accuracy, correlations on test proportions correct were computed within each task. Within the code-todida group all correlations were significant: posttest performance predicted retention test performance (r-square = .78, p = .009), pretest performance predicted posttest performance (r-square = .89, p = .002), and pretest performance predicted retention test performance (r-square = .74, p = .013). The only other significant prediction was within the code-to-letter group for which posttest performance predicted retention test performance (r-square = .83, p = .004).

To examine the ability of posttest reaction time to predict retention test reaction time, and to examine the ability of pretest reaction time to predict posttest and retention test reaction times, correlations on mean logarithmic correct reaction times were computed within each task. Posttest reaction time predicted retention test reaction time within both the code-to-dida group and the dida-to-letter group (r-square = .87, p = .002, and r-square = .79, p = .007, respectively), but not within the code-to-letter group. Pretest reaction time did not significantly predict either posttest or retention test reaction time within any of the task groups.

Error profiles. Across task groups, tests, and number of elements in the stimulus codes, a greater proportion of errors were same-length than would be expected by chance. This was especially notable for the code-to-dida group which made virtually no different-length errors on any test. Within the same-length errors, on 2-element stimulus codes that group made errors at a proportion no different from chance at pretest, made fewer opposite errors (more 1-different errors) than chance at posttest, then returned to levels no different from chance at the retention test. On 3-element codes the proportions of opposite errors were never different from chance, but the group made a higher proportion of 1-different errors and a lower proportion of 2-different errors than chance both on the posttest and on the retention test. The code-to-letter group made a higher proportion of same-length errors than chance although not as dramatically as the code-to-dida group. Within these same-length errors, a higher proportion than chance were opposite errors on the 2-element codes, but on 3-element codes the proportions were not different from chance. Finally, the dida-to-letter group made more same-length errors than chance on all tests on the 3-element codes but not on the 2-element codes. Within these same-length errors, there was a lower proportion of opposite errors than chance on only the posttest for 3-element codes, but again no differences from chance for the 2-element codes. (Analysis of the 2-element errors at posttest was prevented by the low number of dida-to-letter subjects making errors on them.)

The mean proportion of errors that were same-length errors by each task group for both 2-element and 3-element codes is illustrated in Figure 16. Proportions of same-length errors for the 2- and 3-element stimulus codes by each task group were compared to the chance proportions (.27 for 2-element codes, .64 for 3-element codes); because different-length error proportions and same-length error proportions are complementary, comparisons were not carried out on the different-length error proportions.

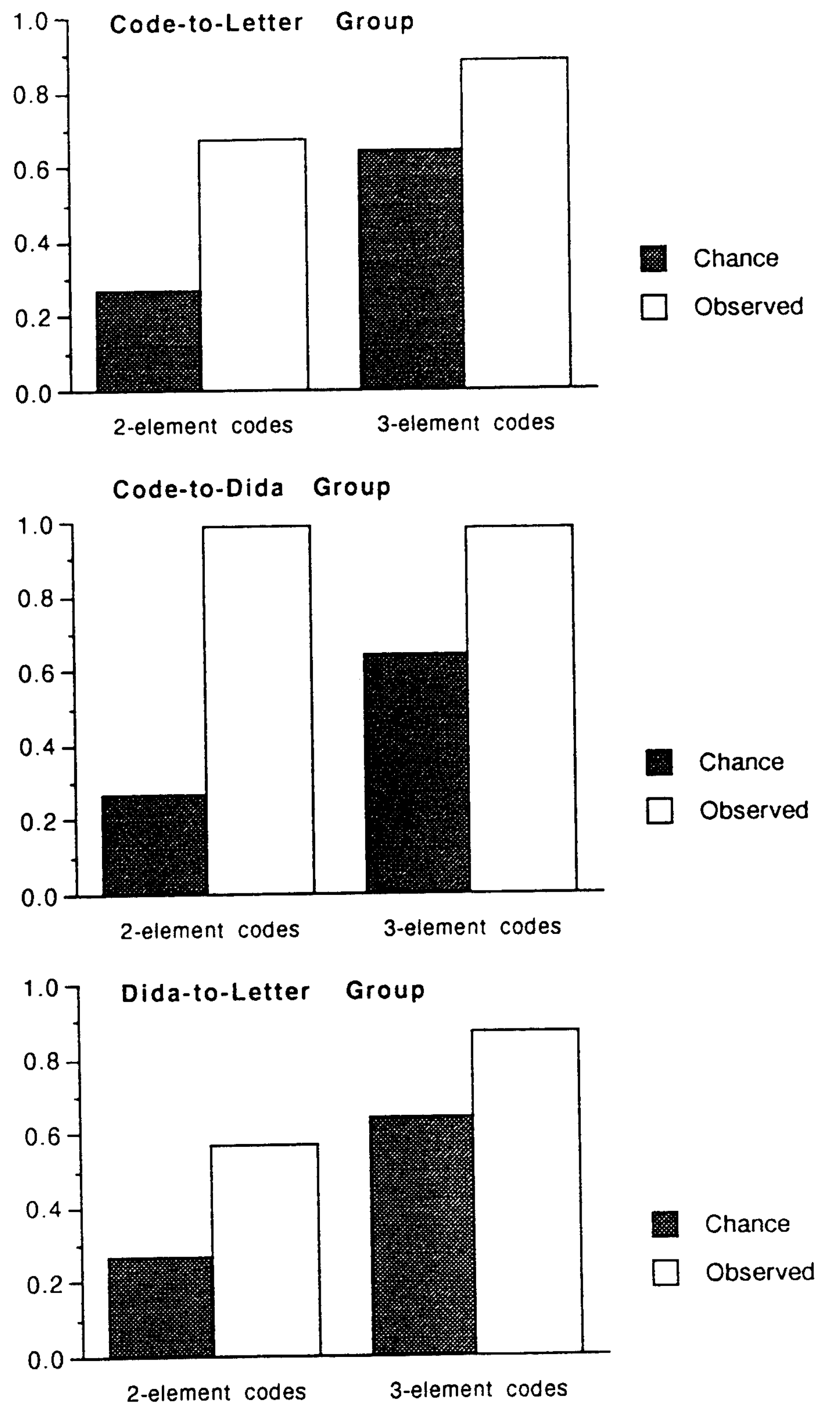


Figure 16. Experiment 3 proportion of errors that were same-length errors on 2- and 3-element stimulus codes.

All three groups made significantly more same-length errors than chance on both the 2-element and 3-element codes: For the code-to-letter group $t(1, 6) = 4.1$, $p < .01$, on 2-element codes, and $t(1, 6) = 8.0$, $p < .001$ on 3-element codes; for the dida-to-letter group $t(1, 6) = 2.5$, $p < .05$, on 2-element codes, and $t(1, 6) = 6.3$, $p < .001$, on 3-element codes. For the code-to-dida group nearly all of the errors were same-length errors, $t(1, 6) = 84.2$, $p < .001$, for 2-element codes, and $t(1, 6) = 28.3$, $p < .001$, for 3-element codes..

For 3-element stimulus codes there are three types of same-length errors: 1-different errors, 2-different errors, and opposite errors. For 2-element codes there are only two types of same-length errors,1-different and opposite errors, because for them 2-different errors are equivalent to opposite errors. Examples of the error types are illustrated in Table 14.

Table 14

Experiment 3 Example Same-Length Error Types

Stimulus Code	1-different errors	2-different errors	opposite errors
· · ·	· · —	· — —	— — —
	· — ·	— · —	
	— · ·	— — ·	
· ·	· —		— —
	— ·		

Analyses of same-length error types were done separately for 2-element and 3-element codes. For 2-element codes the mean proportions of same-length errors that were 1-different and opposite are illustrated in Figure 17; all means were computed using proportions from at least 5 subjects except where noted. For the 2-element codes there were missing values on some tests due to lack of errors (as discussed above and tallied in Table 13); also two subjects in the dida-to-letter group made only different-length errors, one on the pretest and one on the retention test. For these analyses the missing values were left missing. Proportions of opposite errors on each test by each task group were compared to the chance proportion (.33); because opposite error proportions and 1-different error proportions are complementary for 2-element codes, comparisons were not carried out for the 1-different error proportions. On the pretest none of the opposite error proportions were significantly different from chance ($|t| < 2.5$, $p > .070$ for all task groups). The code-to-letter group made significantly more opposite errors than chance on both the posttest and the retention test ($t(1, 4) = 2.8$, $p = .049$ and $t(1, 5) = 3.5$, $p = .018$, respectively). The code-to-dida group made significantly fewer opposite errors than chance on the posttest only ($t(1, 5) = -9.9$, $p < .001$). Note that for the dida-to-letter group the posttest was not analyzed because the number of missing values prevented its analysis.

For 3-element codes the mean proportions of same-length errors that were 1-different, 2-different, and opposite are illustrated in Figure 18. For the 3-element codes there were

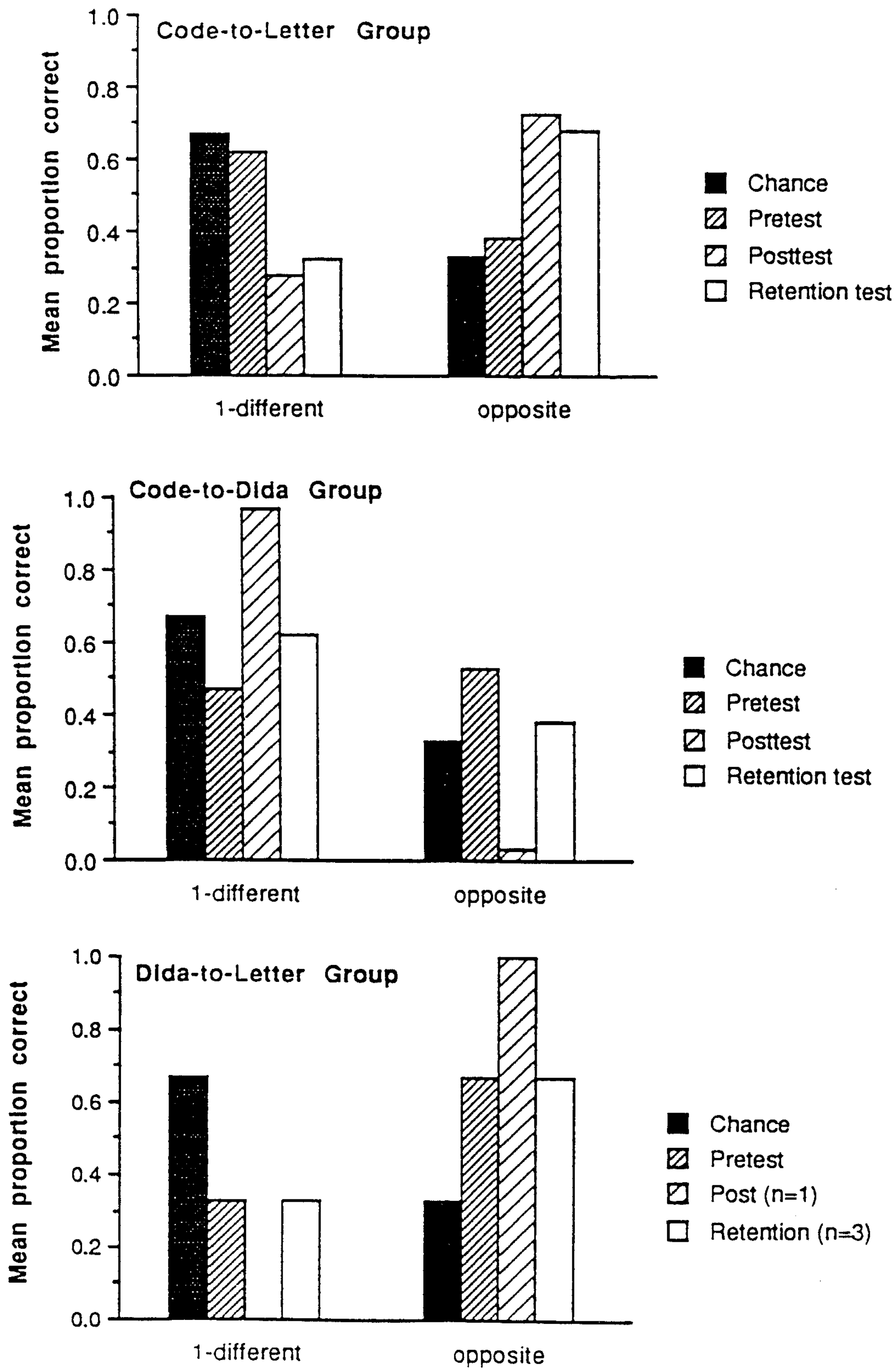


Figure 17. Experiment 3 proportion of same-length errors that were 1-different/opposite errors on 2-element stimulus codes

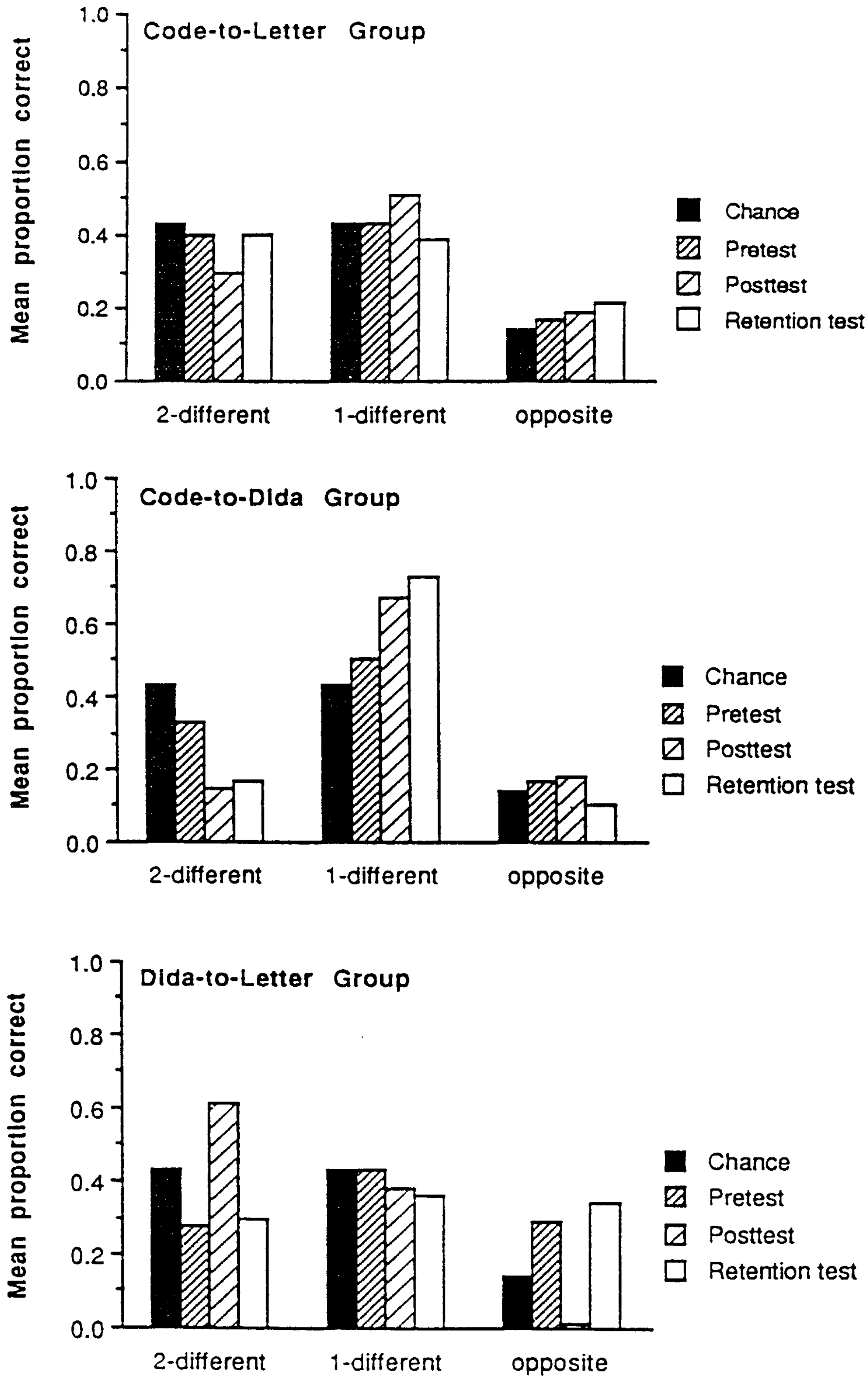


Figure 18. Experiment 3 proportion of same-length errors that were 1-different/2-different/opposite errors on 3-element stimulus codes

missing values only due to lack of errors (as discussed above and tallied in Table 13); for these analyses the missing values were left missing. Proportions of opposite, 1-different, and 2-different errors on each test by each task group were compared to the chance proportions (.14, .43 , and .43 , respectively). On the pretest none of the proportions differed significantly from chance ($/t/ < 2.2$, $p > .070$ for all task groups). The code-to-letter group's error proportions also did not differ significantly from chance on any of the tests ($/t/ < 1.5$, $p > .1$ for all tests). The code-to-dida group made significantly more 1-different and fewer 2-different errors than chance both on the posttest ($t(1, 6) = 3.0$, $p = .023$, and $t(1, 6) = -4.0$, $p = .008$, respectively) and on the retention test ($t(1, 6) = 3.3$, $p = .016$ for 1-different errors, and $t(1, 6) = -3.6$, $p = .011$ for 2-different errors). The dida-to-letter group made significantly fewer opposite errors than chance on the posttest only, $t(1, 4) = -8.1$, $p = .001$.

Summary

Observed accuracy performance lent some support to the a priori decomposition of the complete task into the code-to-dida and dida-to-letter subtasks, although this decomposition is best understood as an approximation to the whole task. Task did not affect accuracy although it affected reaction times with the code-to-dida consistently faster than the other groups. Overall, reaction times and accuracy both suggested that 2-element codes are easier than 3-element codes and that performance improved with practice then worsened over retention. These effects were influenced by task; most noticeably, the code-to-dida group showed consistency across tests both in terms of accuracy and of reaction times with no loss over retention.The effect of letter difficulty on Morse reception is not peculiar to only one of the subtasks. Subjects in all three tasks showed a performance advantage for the easy subset of stimuli, in terms of both accuracy and reaction time, although the reaction time advantage for the code-to-dida group was smaller than for the other groups.

Within the code-to-dida group, accuracy was consistently predictable throughout the experiment, and reaction time on the posttest predicted reaction time on the retention test. Posttest performance predicted retention test performance within the code-to-letter group for accuracy only and within the dida-to-letter group for reaction time only.

Analysis of the error profiles revealed that across task groups, tests, and number of elements in the stimulus codes, a greater proportion of errors were same-length than would be expected by chance. This was especially notable for the code-to-dida group which made virtually no different-length errors on any test. Within the same-length errors, on 2-element stimulus codes that group made errors at a proportion no different from chance at pretest, made fewer opposite errors (more 1-different errors) than chance at posttest, then returned to levels no different from chance at the retention test. On 3-element codes the proportions of opposite errors were never different from chance, but the group made a higher proportion of 1-different errors and a lower proportion of 2-different errors than chance both on the posttest and on the retention test. The code-to-letter group made a higher proportion of same-length errors than chance although not as dramatically as the code-to-dida group. Within these same-length errors, a higher proportion than chance were opposite errors on the 2-element codes, but on 3-element codes the proportions were not different from chance. Finally, the dida-to-letter group made more same-length errors than chance on all tests on the 3-element codes but not on the 2-element codes. Within these same-length errors, there was a lower proportion of opposite errors than chance on only the posttest for 3-element codes, but again no differences from chance for the 2-element codes. (Analysis of the 2-element errors at posttest was prevented by the low number of dida-to-letter subjects making errors on them.)

CHAPTER V

GENERAL DISCUSSION

Experiment 1, 2

Acquisition. Difficult-first training did not lend a lasting advantage during acquisition either in terms of accuracy or in terms of reaction time. This contrasts with Pellegrino et al.'s (1991) finding that difficult-first training led to faster performance after further training on all stimuli. The many differences between Pellegrino et al.'s visual discrimination experiment and Experiments 1 and 2 are noted in the introduction to Experiment 1. Probably the most crucial difference, however, lies in the difficulty of the easy items. In the visual discrimination experiment all items were relatively easy as evidenced by the error rates' being "generally low" (Pellegrino et al., p. 783); in this experiment even the easy items were relatively challenging, with high initial error rates (about 50% on the Session 1 pretests). Pellegrino et al.'s explanation for their effect -- that the easy items allowed subjects to use a loose net of discriminations depending on only a general strategy whereas the difficult items encouraged greater specificity of the stimulus encoding network -- depends on the easy items' being easy enough for a loose net to be successful. I propose that in these Morse code experiments, the easy items themselves were difficult enough to require that subjects adopt item-specific strategies that also stood them in good stead on the difficult items as well.

Retention. There was a large retention advantage on easy letters for the easy-first group in Experiment 2 and a trend for a retention disadvantage for the difficult-first group in Experiment 1. Such an effect could be explained by displaced rehearsal leading to greater decline for the difficult-first group on easy letters than for the easy-first group. To paraphrase Slamecka and Katsaiti (1987) in their discussion of the generation effect, a within-list variable allows subjects to rob Peter to pay Paul; that is, they can concentrate on one subset of the trials to the detriment of the other subset. Although during the initial session of training the easy-first group mastered the easy letters, the difficult-first group did not master the difficult letters. It would be possible therefore that during further training the difficult-first group neglected the new easy codes to concentrate still on the difficult codes. However, the planned Session 3 analyses did not show the interaction of training condition and letter difficulty which would have been expected if the difficult-first group were neglecting the easy letters and displacing rehearsal to the difficult letters during acquisition.

Overlearning of the easy letters by the easy-first group does explain the easy-first advantage at retention. Overlearning occurs when subjects train longer than is required to reach a criterion. If ceiling performance is taken as a post hoc criterion, then these experiments do show some overlearning.The easy-first group reached ceiling performance on easy letters after the first session of training but continued to receive training on those items for two more sessions, hence overlearning the easy letters. The all-first group of Experiment 1 reached ceiling on the easy items after two sessions but continued training on easy letters for one more session, overlearning the easy items but to a lesser degree than did the easy-first group. Whether overlearning leads to an actual lower forgetting rate is not clear (Ericsson & Crutcher, 1988), but in this case the overlearning would cause an apparent lower forgetting rate. Because the easy-first group performed at ceiling on the easy items of the Session 3 posttest their easy-first score on that test is not representative of their true skill level.

In Experiment 2 the accuracy levels of the difficult-first group on easy and on difficult letters and of the easy-first group on difficult letters all drop more than does that of the easy-first group on easy letters. Experiment 1 holds the key to illuminating this effect of ceiling performance on retention level: In Experiment 1 the accuracy levels of the difficult-first group on easy and on difficult letters and of the easy-first group on difficult letters, as well as of the all-first group on the difficult letters, all drop more than does that of the easy-first group on easy letters; the all-first group's drop on easy-first letters, however, is intermediate. This is key because on the easy letters the group that had shown ceiling performance the longest (the easy-first group) showed the best retention, followed by the group that had reached ceiling after two sessions (the all-first group), followed finally by the group that never attained ceiling performance (the difficult-first group). Because number of sessions at ceiling is a measure of overlearning which can be a measure of degree of original learning, these results do not argue for a special case for the easy-first group on easy letters. In Experiment 2 the easy-first group reached ceiling on easy letters but the other proportions correct were again under .80. Because the skill level of the easy-first group on easy letters was higher than their scores suggest, their retention performance could represent a drop in skill level commensurate with that for the difficult group.

Experiment 3

Subtasks. From the analyses in Experiment 3 it is clear that the code-to-dida subtask of Morse code reception is a qualitatively different task from either the dida-to-letter subtask or the complete code-to-dida task. Reaction times were much faster for the code-to-dida group than for the others throughout the experiment. Predictive validities were much greater for that group than for the others. Improvement after training for the code-to-dida group was slighter than for the other groups, and whereas both other groups experienced loss over retention the code-to-dida group showed none. The code-to-dida skill is a largely procedural one, related to the prototypic procedural task of typing; the dida-to-letter task involves a largely declarative skill related to the prototypic declarative task of learning paired associates, and the code-to-letter task includes this component as well. Therefore the procedural reinstatement advantage for procedural skills holds in this case, lending the code-to-dida group a retention advantage.

Sources of error in Morse code reception. The code-to-dida group was also qualitatively different from the other groups in that it made virtually no different-length errors on any of the tests; this finding agrees with Shepard's (1963) interpretation of Rothkopf's (1957) Morse code discrimination study that confusion errors tend to have the same number of elements as the stimulus code. Because in Rothkopf's discrimination study subjects dealt only with the codes without associating them with letters, Rothkopf's task is most closely related to the code-to-dida subtask. Shepard's interpretation that errors in Rothkopf's study also tended to have the same predominance of dots and dashes (which would imply fewer opposite errors because those can completely switch the predominance) partly agrees with code-to-dida performance in that fewer opposite errors than chance were committed on 2-element errors after training, but not so on the 3-element codes.

Shepard's analysis of novice code reception performance suggested that confusion errors should again be of the same length as the stimulus code and that they should have the same degree of heterogeneity. This implies that there should be more same-length errors and more opposite errors than chance (because opposite errors preserve the degree of heterogeneity). Indeed the code-to-letter group did make more same-length errors than chance and more opposite errors than chance on the 2-element codes after training; however there were not more opposite errors than chance on 3-element codes. The source of the higher-than-chance level of opposite errors on 2-element codes in the full Morse code reception task is not clear from this experiment because the dida-to-letter group did not show more opposite errors than chance after training (and indeed showed fewer than chance on 3-element codes) nor did the code-to-dida group; it must be remembered however that the dida-to-letter group's posttest errors on 2-element codes could not be analyzed. The code-to-dida subtask findings cast light on the difference between Shepard's (1963) interpretation of Morse code discrimination findings and his interpretation of novice code reception findings. The agreement suggests that the importance of dot/dash predominance (rather than of element heterogeneity) for code discriminations is due to the perceptual properties of the codes rather than to the discrimination task itself.

Predictive Validities

Table15 summarizes the findings concerning predictive validity for all three experiments (note that the within-group results from Experiment 1 were not included in the table because the number of subjects in each group was so small). For each measure on each experiment the table includes a column for whether or not the posttest accuracy predicted retention test accuracy, a column for whether or not pretest accuracy predicted retention test accuracy, and columns for the same questions about reaction times. For Experiments 1 and 2 the Session 1 pretest is considered the "pretest", the Session 3 posttest is considered the "posttest" and the Retention Session pretest is considered the "retention test". Two further columns are included in the table; these are for whether or not group membership predicted retention test performance above and beyond the prediction by posttest performance.

Table 15

Significant Predictors of Retention Level (✓) and Predictors that were Tested and Not Found Significant(O)

	Posttest accuracy	Pretest accuracy	Posttest RT	Pretest RT	Group accuracy	Group RT
Exp. 1 Across groups						
overall	✓				O	
easy subset	✓				O	
difficult subset	✓				O	
Exp 2 Across groups						
overall	✓	O	✓	✓	✓	✓
easy subset	✓	O	✓	✓	✓	✓
difficult subset	✓	✓	✓	✓	O	O
Exp. 2 Within easy-1st group						
overall	✓	O	✓	O		
easy subset	✓	O	✓	O		
difficult subset	✓	✓	✓	O		
Exp. 2 Within difficult-1st gp.						
overall	O	✓	O	✓		
easy subset	O	O	O			
difficult subset	✓	O	✓	✓		
Exp 3						
code-to-ltr group	✓	O	O	O		
code-to-dida group	✓	✓	✓	O		
dida-to-ltr group	O	O	✓	O		

From Table 15 it is clear that in most but not all cases posttest performance did indeed predict retention performance both in terms of accuracy and in terms of reaction time. This predictive validity cannot be explained completely in terms of untrained aptitude for the tasks because pretest performance did not predict retention test performance on many of these tasks. However, the predictive ability of group membership is important to keep in mind because it indicates that posttest performance is not the only predictor of retention level. The initial training set of the Experiment 2 subjects predicted their retention test performance even when posttest performance had already been considered.

Conclusions

These experiments have demonstrated that there is no advantage for difficult subset initial training on the task of Morse code reception. It is interesting that in Experiment 1 whole training (all-first) yielded very good results. All-first subjects showed performance on the difficult subset similar to that of the difficult-first group and performance on the easy subset that approached that of the easy-first group. This finding seems impressive because it involved learning 12 discriminations and identifications at the same time. However, Keller (1943) and Spragg (1943) both demonstrated that after 12 whole set sessions of training, sessions that were not quite twice as long as those of Experiments 1 and 2, on a whole set of the entire alphabet, most subjects could achieve 95 percent accuracy (in Keller's study the whole set consisted of the alphabet plus numbers one through nine).

In Experiment 3 error patterns for the code-to-dida subtask, the dida-to-letter subtask, and the code-to-letter whole task agreed with earlier error studies that found most errors in code discrimination and novice code reception had the same number of elements as the stimulus code but gave mixed agreement to previous findings of the importance of dot/dash predominance in discrimination and of element heterogeneity in code reception.

The stability of individual differences for the code-to-dida group, indicated by their consistently significant predictive validities, suggests that the subtask of segmenting codes into their element patterns could be a limiting factor for some students of Morse code reception. This finding raises three issues for future research. The first is whether the individual differences in code-to-dida skill level remain stable over longer periods of training, periods of training similar to those of Morse code students attempting to master the entire set of codes. Second, are there strategies in either code-to-dida or whole-task (code-to-letter) performance that would allow individuals to overcome low code-segmenting abilities? Finally, given the stability of individual differences in segmenting codes, what would be the effectiveness of part training of the subtasks on transfer to the complete task of Morse code reception?

Finally, in Experiment 3 it was demonstrated that the code-to-dida subtask of Morse code reception is qualitatively different from the other subtask or the whole task. Furthermore, the ease of procedural reinstatement for this subtask could explain its higher retention than for the other tasks. In their discussion of procedural reinstatement, Healy et al. (in press) discount three other possible explanations for their findings of higher retention for some skills than for others. The high levels of retention for those skills could not be accounted for by advantages for automatic skills, by advantages on implicit memory measures, or by advantages for motor skills. Two of these three alternative explanations can be discounted for the code-to-dida group findings as well. The code-to-dida skill, with a posttest mean accuracy proportion of .70, could not be automatic, and the explicit rather than implicit memory measure of recall accuracy showed high retention for the code-to-dida group. However, the code-to-dida skill was a perceptual-motor skill, whereas the other Experiment 3 tasks were not.

Procedural reinstatement predicts high levels of retention for tasks that use the same context at test as at learning and/or that are procedural rather than declarative tasks. Each of the three tasks in Experiment 3 used identical contexts for training and testing (with the exception of feedback), so contexts cannot explain the results. Procedural reinstatement predicts that procedural skills will be better retained than declarative information. The three features of procedural versus declarative memory should therefore distinguish between tasks that are well-retained and those that are not; these three features are possible partial acquisition versus all-or-none acquisition, gradual acquisition by performing versus sudden acquisition possibly by being told, and expression of the knowledge directly versus flexible/verbal expression (not included here is Squire's,1987, addition that procedural skills can be modality bound but declarative skills are not, because Healy et al.'s, in press, mental multiplication study found high retention even when the modality of response was switched from typing to oral). Of these three, the first does not distinguish between the three tasks because in all three tasks trials were judged simply correct or incorrect; there was no "partially correct" (and in the error analysis all three tasks had errors classified as 1-different, 2-different, and so on; all three had degrees of error). The predictive ability of the second feature of declarative versus procedural memory is unclear because the meaning of "gradual" is unclear, but the code-to-dida group did appear to have a less steep learning slope than the other groups; however, for all three groups practice consisted of performing the tasks rather than being told about them. Regarding the third feature, because verbal protocols were not collected on Experiment 3, it cannot be categorically stated that the code-to-dida group could not verbally express their knowledge.

Application of the three features of procedural versus declarative knowledge to other tasks is equally spotty. Of the six tasks discussed by Healy et al. (in press) only one, data-entry, could be said to have been partially possessed, because its primary performance measure was reaction time which is not an all-or-none measure even at the single-trial level. Comparing the six tasks on whether or not they were gradually acquired by performance, one must again wrestle with the meaning of gradual. However, determining whether or not a task was acquired by performance is relatively straightforward, and the tasks that explicitly were not so acquired -- memory for calculation results and memory for who, what, and when information regarding course schedules -- yielded low retention. Finally, the third feature, the question of verbal expression, is unclear. What does it mean to be unable to verbalize a skill? If subjects find it impossible to verbalize a skill, as in Anderson's (1976) example of a well-practiced driver's inability to say whether or not to take one's foot off the accelerator when shifting, then that skill is procedural, but the other argument, that any ability to verbalize a skill means it is a declarative skill does not support retention level findings. For example, is the potential protocol, "I saw the number '538' and entered it" (for the high retention data entry task) less of a verbalization than the potential protocol, "I saw the word 'doronico' and said, 'leopard'" (for the low retention vocabulary task)?

A priori prediction of retention level is an important topic for further research. Procedural reinstatement can readily predict low retention for tasks that require different contexts at test than during practice. However, fine-tuning the operational definitions of the features of procedural versus declarative information as regards retention predictions, and determining whether these features are necessary or sufficient for procedural reinstatement, must still be accomplished.

REFERENCES

Anderson, J. R. Language, memory and thought. Hillsdale, NJ: Lawrence Erlbaum Associates, 1976.

Annett, J, & Kay, H. (1956). Skilled performance. Occupational Psychology, 30, 112-117.

Ardila, A. (1987) Aphasia for Morse code: A comment on Wyler and Ray(1986). Brain and Language, 30, 363-366.

Bryan, W. L., & Harter, N. (1899). Studies on the telegraphic language. The acquisition of a hierarchy of habits. Psychological Review, 6, 346-375.

Campbell, W. J., & Allen, G. L. (1987). Experimentally assessed learning rate as a determinant of performance under self-paced and instructor-paced instruction. Journal of Psychology, 122, 69-78.

Cormier, P., Tomlinson-Keasey, C., & Geary, D. C. (1988). Gender and hemispheric specialization differences in the learning of Morse code letters. Bulletin of the Psychonomic Society, 26, 399-402.

Ericsson, K. A., & Crutcher, R. J. (1988). Long-term retention of sequentially organized information, knowledge, and skills: An empirical review (Technical Report #88-14). Boulder, Colorado: University of Colorado at Boulder, Institute of Cognitive Science

Farr, M. J. (1987). The long-term retention of knowledge and skills: A cognitive and instructional perspective. New York: Springer-Verlag.

Gibson, E. J. (1969). Principles of perceptual learning and development. New York: Meredith Corporation.

Healy, A. F., Fendrich, D. W., Crutcher, R. J., Wittman, W. T., Gesi, A. T., Ericsson, K. A, & Bourne, L. E., Jr. (in press). The long-term retention of skills. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), From learning processes to cognitive processes: Essays in honor of William K. Estes, Volume 2 (pp. 87-118). Hillsdale, NJ: Erlbaum.

Healy, A. F., Fendrich, D. W., & Proctor, J. D. (1990). Acquisition and retention of a letter-detection skill. Journal of Experiment Psychology: Learning, Memory, and Cognition, 16, 270-281.

Highland, R. W., & Fleishman, E. A. (1958). An empirical classification of error patterns in receiving Morse code. Journal of Applied Psychology, 42, 112-119.

Keller, F. S. (1943). Studies in international Morse code. 1. A new method of teaching code reception. Journal of Applied Psychology, 27, 407-415.

Keller, F.S., & Taubman, R.E. (1943). Studies in international Morse code. 2. Errors made in code reception. Journal of Applied Psychology, 27, 504-509.

Klapp, S. T., & Wyatt, E. P. (1976). Motor programming within a sequence of responses. Journal of Motor Behavior, 8, 29-26.

Kolers, P. A., & Roediger, H. L. (1984). Procedures of mind. Journal of Verbal Learning and Verbal Behavior, 23, 425-449.

Mane, A. M., Adams, J. A., & Donchin, E. (1989). Adaptive and part-whole training in the acquisition of a complex perceptual-motor skill. Acta Psychologica, 71, 179-196.

Morsh, J. E., & Stannard, A. F. B. (1947). Studies in international Morse code. III. Retention of training by Morse code operators. Canadian Journal of Psychology, 1, 141-144

Naylor, J. C., & Briggs, G. E. (1963). Effects of task complexity and task organization on the relative efficiency of part and whole training methods. Journal of Experimental Psychology, 65, 217-232.

Newby, T. J., Cook, J. A., & Merrill, P. F. (1988). Visual mediational instruction: Reducing interference within visual and aural multiple-discrimination tasks. Journal of Educational Psychology, 80, 40-45.

Newell, K. M., Carlton, M. J., Fisher, A. T., & Rutter, B. G. (1989). Whole-part training strategies for learning the response dynamics of microprocessor driven simulators. Acta Psychologica, 71, 197-216.

Pellegrino, J.W., Doane, S.M., Fischer, S.C., & Alderton, D. (1991). Stimulus complexity effects in visual comparisons: the effects of practice and learning context. Journal of Experimental Psychology: Human Perception and Performance, 17, 781-791.

Plotkin, L. (1943). Stimulus generalization in Morse code learning. Archives of Psychology. New York, 40, No. 287.

Rothkopf, E.Z. (1957). A measure of stimulus similarity and errors in some paired-associate learning tasks. Journal of Experimental Psychology, 53, 94-101.

Rothkopf, E. Z. (1958). Stimulus similarity and sequence of stimulus presentation in paired-associate learning. Journal of Experimental Psychology, 56, 114-122.

Seashore, H., & Kurtz, A.K. (1944). Analysis of errors in copying code (Contract No. OEMsr-830). Washington, DC: Office of Scientific Research and Development, Report No. 4010. (U.S. Department of Commerce Publ. Bd. No. 12170; ASTIA Document No. ATI-35548)

Seymour, W. D. (1954). Experiments on the acquisition of industrial skills. Occupational Psychology, 28, 77-89.

Shepard, R.N. (1963). Analysis of proximities as a technique for the study of information processing in man. Human Factors, 5, 33-48.

Slamecka, N.J., & Katsaiti, L.T. (1987). The generation effect as an artifact of selective displaced rehearsal. Journal of Memory and Language, 26, 589-607.

Squire, L. R. (1987). Memory and Brain. New York: Oxford University Press.

Spragg, S. D. S. (1943). The relative difficulty of Morse code alphabet characters learned by the whole method. Journal of Experimental Psychology, 33, 108-114.

Stammers, R. B. (1982). Part and whole practice in training for procedural tasks. Human Learning, 1, 185-207.

Tulving, E. (1985). How many memory systems are there? American Psychologist, 40, 385-398.

Wyler, A. R., & Ray, M. W. (1986). Aphasia for Morse code. Brain and Language, 27, 195-198.