

Speech-like coding of pictures in short-term memory

DIANE J. SCHIANO and MICHAEL J. WATKINS
Princeton University, Princeton, New Jersey 08544

Memory span for pictures of common objects and for the names of these objects was examined as a function of three speech-related variables. Both picture span and name span were found to be influenced by the phonological similarity (Experiment 1) and the length (Experiment 2) of the names, as well as by the subject's engaging in "irrelevant" vocalization during item presentation (Experiment 3). Moreover, for each variable the effect was in the same direction and of comparable magnitude for the two types of items. Experiments 4-6 replicated these findings with the procedure modified such that the retention of order information was not required. It is concluded that under the present conditions, there is a substantial functional equivalence between short-term memory for readily nameable pictures and for words and that this equivalence may be thought of as due to mediation by a common, "speech-like" code.

In terms of the familiar metaphor that memory entails the storage of coded information, this paper is concerned with the relation between the coded information underlying the immediate recall of pictures and of words. More particularly, it reports a series of experiments designed to see whether, in a memory-span situation, the codes for pictures and for words are of the same sort or are qualitatively different. For convenience, we will refer to these alternatives as the same-code hypothesis and the different-code hypothesis.

In most of the research on short-term memory, the to-be-remembered material has been verbal, and several findings have been taken as suggesting that the memory code for verbal material is speech-like. Three such findings have been particularly important: (1) When the to-be-remembered or "memory" items all sound somewhat alike, recall is poorer than when they all sound quite different (Baddeley, 1966; Conrad, 1963, 1964; Estes, 1973); (2) lists of lengthy words are recalled less well than are lists of short words (Baddeley, Thomson, & Buchanan, 1975); and (3) recall is poorer when the subject is required to chant other, irrelevant items during presentation of the memory items (Levy, 1971; Murray, 1968; Richardson & Baddeley, 1975). Each of these findings fits in well with the notion that with verbal items retention is mediated by a speech-like code. Thus, it seems reasonable to suppose speech-like codes to be more confusable for items that sound alike, to

contain more information and so consume more capacity for items that are polysyllabic, and to be less well formed or more interfered with when the subject has to chant other sounds.

These effects have been found not only when the verbal items are presented auditorily, but also when they are presented visually. This fact is of some importance, for it demonstrates a certain functional equivalence in the short-term remembering of spoken and written words. More particularly, it suggests the retention of both spoken and written words to be mediated by memory codes that have speech-like characteristics.

The question arises of whether this functional equivalence extends to readily nameable pictures. That is, is short-term memory for pictures also susceptible to these speech-related variables? Some indirect support for this possibility is suggested by the finding that concurrent performance of an attention-demanding auditory shadowing task may impair memory for pictures as well as for words (Rowe & Rogers, 1975). More directly, there is some evidence that picture memory is reduced when the names of the pictures are phonologically similar to each other (Conrad, 1971, 1972). Perhaps, therefore, we should think of short-term memory for pictures as well as for words as being mediated by speech-like codes. The purpose of the experiments reported here was to pursue this possibility by seeing whether name length and "irrelevant" vocalization, as well as phonological similarity, affect short-term memory for pictures. Whereas the different-code hypothesis makes no prediction about how these speech-related variables should affect memory for pictures, the same-code hypothesis predicts that they will affect picture memory in just the same way that they affect word memory.

This research was supported by the National Institute of Mental Health Grant MH31674 to Michael J. Watkins. The authors are grateful to Olga C. Watkins for criticism of an earlier draft of the paper. Requests for reprints may be sent to Michael J. Watkins, who is now at the Department of Psychology, Rice University, Houston, Texas 77001.

EXPERIMENTS 1-3: ORDERED RECALL OF PICTURES AND WORDS

The first three experiments used an ordered reconstruction procedure to estimate memory span under four conditions: specifically, those given by combining two modes of item presentation (picture and word) with two levels of a speech-related variable. This variable was phonological similarity in Experiment 1, word (name) length in Experiment 2, and irrelevant vocalization in Experiment 3. Since the experiments were all very similar, their common methodology is described first, and then details specific to the individual experiments are noted.

Method

Materials. Each stimulus item was presented on a separate 3 x 5 in. (7.6 x 12.6 cm) card. The words were printed in letters .75 in. (1.9 cm) high. The pictures were simple black-and-white line drawings of familiar objects, similar to those found in dictionaries. The items in a given word condition were the names of the objects in the corresponding picture condition. There could therefore be no confound of mode of presentation with the nominal identity of the stimulus items. Duplicate sets of cards were prepared for the subjects' response task.

Subjects. A total of 41 young adults from the Princeton, New Jersey, community participated as paid volunteers. Each subject served in only one experiment.

Design. The experiments were designed to compare memory spans for the four conditions. Subjects were tested individually. They were presented a sequence of 52 lists of items, and their task was to remember the items in their order of presentation. The 52 lists comprised 13 in each of the four conditions. Each successive block of four lists included one list from each condition, with ordering of conditions separately randomized for each block; the only constraint on the randomization was that the last condition in one block was never repeated as the first condition in the next.

An estimate of memory span was made for each of the four conditions, using the "up-and-down" method. Thus, the number of items in a given list was one greater than that of the last list in that condition if the latter had been recalled correctly, and one less if not. For the first block, all four lists comprised four items. For the second block, list length for a given condition was three or five items, depending on performance in that condition in the first block, and so forth for successive blocks.

A general problem in comparing verbal and pictorial memory concerns the nature of the response requirements. In the present case, it was considered desirable to use a common mode of response and, moreover, one that is inherently neither verbal nor pictorial. To this end, an ordered reconstruction task was used: For each condition, the subject was given a duplicate deck of cards, and he or she responded by selecting those presented and arranging them in their order of presentation.

Procedure. The subject was seated at a table across from the experimenter and was informed about the experiment in general and about the four conditions in particular. To increase the probability that any implicit naming by the subject involved the intended names, the experimenter showed and named each picture in turn. The presentation and duplicate decks for each condition were shuffled and placed face down before the experimenter and the subject, respectively. For each trial, the experimenter took the appropriate presentation deck and, working behind a screen, constructed the list by randomly selecting the required number of cards and adding a blank end-of-list card. The subject was informed of the condition of the to-be-presented list, and he or she placed the corresponding duplicate deck at the ready, although still face down. The experimenter then pre-

sented the list by placing the cards one on top of the other at an even rate. On seeing the blank card, the subject turned over the duplicate deck, selected the cards that he or she thought had been presented, and attempted to lay them out on the table with a left-to-right ordering representing temporal order of presentation. A change in placement of a card already positioned was allowed, and no time limit was imposed. Care was taken to ensure that the subject was fully aware that credit was gained only for items placed in their exact presentation positions. He or she was told after each trial whether the list had been reconstructed correctly; if it had not, the experimenter displayed the cards as they had been presented. The next trial began after a short break in which the experimenter prepared the next list and the subject gathered the cards that he or she had just used and then shuffled the deck. Note that throughout the sequence of trials neither the experimenter nor the subject overtly named the items.

Experiment 1: Phonological similarity. The four conditions in Experiment 1 corresponded to four types of items: (1) phonologically dissimilar words, (2) pictures with phonologically dissimilar names, (3) phonologically similar words, and (4) pictures with phonologically similar names. The words were the names of the pictures, and all were monosyllabic.

There were 13 subjects, each of whom worked with a different set of items. Counterbalancing individual items with respect to similarity condition was achieved by selecting items that could be arranged into a 12 by 13 matrix, in which the 12 items within a column all shared the same vowel sound (e.g., crib, pig, fish, whip, etc.) and the 13 items within a row had a different vowel sound (e.g., crib, hat, pipe, key, etc.). For Subjects 1-12, the words for the similar condition were those in Columns 1-12, respectively, and those for the dissimilar condition were those in Rows 1-12, respectively. The items common to a row and column were used only in the similar condition, so that for each subject, there were 12 words in the similar condition and a different set of 12 in the dissimilar condition. The 12 "common" words, which fell on what is approximately a diagonal of the matrix, served as the dissimilar words for Subject 13; the words in Column 13 served as this subject's similar words. In this way, each subject saw a unique set of items, but across subjects, every word served once in the similar and once in the dissimilar condition. The items for the picture conditions corresponded directly to those of the word conditions. Thus, for any given subject, the words for the word-similar condition were the names of the pictures of the picture-similar condition, and likewise for the dissimilar conditions. The cards were presented at a rate of 1 card/sec.

Experiment 2: Word (name) length. Like those of the first experiment, the four conditions in Experiment 2 corresponded to four types of items: (1) short (one-syllable) words, (2) pictures with short names, (3) lengthy (three- to five-syllable) words, and (4) pictures with lengthy names. For a given subject, there were 10 cards for each condition, and as before, the items for each of the word conditions were the names of the objects depicted by the pictures in the corresponding picture conditions. To ensure that any differences between the two picture conditions could be safely attributed to their names and not to unintended differences among the pictures themselves, the pictures were selected and drawn in such a way that each could be given both a short name (e.g., church, moth, suit) and a lengthy name (e.g., cathedral, butterfly, tuxedo). The assignment of each picture to the short or lengthy condition was counterbalanced across subjects. Items were presented at a 2-sec rate. Sixteen subjects participated.

Experiment 3: Irrelevant vocalization. In this experiment, there were only two types of items: pictures and their names. The four conditions were given by combining these two types of items with two study conditions, namely, with and without irrelevant vocalization. For the irrelevant vocalization condition, subjects counted aloud from one to three during the presentation of each item. The items were presented at a 1-sec rate, and practice in the vocalization task was given prior to testing. To

keep procedural details as similar as possible to those of the first two experiments, a separate deck of cards was used for each condition. Thus, the picture and word decks within each study condition depicted the same set of objects, but between study conditions, two different sets were depicted. The memory items were the pictures and short words used in Experiment 2; there were 10 cards in each deck. Assignment of the two pairs of decks to study conditions was counterbalanced across two groups of six subjects. There were therefore 12 subjects in all.

Results and Discussion

Since the up-and-down method continually adjusts list length to keep the subject's rate of perfect recall at about the 50% level, memory span is estimated by the mean length of the lists presented. In making this estimate, the first few lists are best ignored, as they reflect the arbitrary length of the first list. In the present case, each subject's span for each condition was estimated by averaging the lengths of Lists 5-14. (Although only 13 lists were presented, the length of List 14 was given by the length and outcome of List 13.) The mean spans for the four conditions in each of the first three experiments are shown in Table 1.

In all three experiments, the speech-related variable affected mean span for words in the expected direction. That is, memory span was smaller with phonologically similar words, with lengthy words, and with irrelevant vocalization [for Experiment 1 (phonological similarity), $t(12) = 5.86$; for Experiment 2 (word length), $t(15) = 4.68$; for Experiment 3 (irrelevant vocalization), $t(11) = 9.53$; $p < .01$ in each case]. Of more interest is the finding that in each experiment, the speech-related variable affected picture span in the same way that it affected word span. Again, this effect was significant [for Experiment 1, $t(12) = 5.33$; for Experiment 2, $t(15) = 2.76$; for Experiment 3, $t(11) = 8.87$; $p < .01$ in each case]. Moreover, there was no reliable evidence that any of the three variables affected word span more than picture span. Although the effects were greater for words in Experiments 1 and 2, they were not significantly so [$t(12) = 1.25$, $p > .10$, and $t(15) = .73$, $p > .10$, respectively]. In Experiment 3, words were slightly less affected than were pictures.

It is also worth noting that in each of these experiments there was a small overall advantage for pictures over words. The difference was significant in Experiment 2 [$t(15) = 2.91$, $p < .01$], but not in Experiment 1 [$t(12) = .65$] or Experiment 3 [$t(11) = .25$] ($p > .10$ in each case).

The overall pattern of results for the first three experiments is consistent with the same-code hypothesis. More particularly, it is consistent with the notion that in a memory-span situation, effective coding for pictures as well as for words is speech-like. Certainly, the finding that the various speech-related variables affected the pictures no less than the words is contrary to any simple formulation of the different-code view.

One way to interpret these results, as well as Conrad's (1971, 1972) earlier results concerning phonological similarity, is in terms of what is known as dual-code

Table 1
Estimates of Memory Span in Experiments 1-3 (Ordered Recall)

Experiment	Variable	Condition	Words	Pictures
1	Phonological Similarity	Dissimilar	6.08	6.04
		Similar	4.70	4.96
2	Word (Name) Length	Short	6.49	6.81
		Lengthy	5.60	6.16
3	Irrelevant Vocalization	Silent	5.85	5.92
		Chanting	3.97	3.98

theory. This theory, formulated and developed by Paivio (e.g., Paivio, 1971), is a sophisticated composite of the same-code and different-code hypotheses. It assumes that an encounter with either a picture or a word may give rise to both an "image" and a "verbal" code, but that an image code is formed more readily for a picture and a verbal code is formed more readily for a word. It also assumes that retention of order information is mediated by the verbal code (e.g., Paivio & Csapo, 1969; Paivio, Philipchalk, & Rowe, 1975). It is this latter assumption that enables the theory to account for our findings, for by requiring subjects to arrange the items in their temporal order of presentation, we may have rendered useless the image code and assured the exclusive use of the verbal code.

If this interpretation is appropriate, then a somewhat different set of results would be expected if order information were not required. It seems reasonable to suppose that in this case the use of an image code would—if not for words, then surely for pictures—mask any damaging effects that an experimental manipulation might have on the verbal codes. On the other hand, a simple version of the same-code hypothesis has nothing to say about the retention of order information, and so there would be no reason to expect that freeing the subject from the need to remember order information would change the pattern of results. It was the purpose of Experiments 4-6 to put these contrasting predictions to test.

EXPERIMENTS 4-6: FREE RECALL OF PICTURES AND WORDS

Method

The designs and procedures of Experiments 4-6 followed closely those of Experiments 1-3. Thus, we investigated the effects of phonological similarity, word (name) length, and irrelevant vocalization on memory for both words and pictures in Experiments 4, 5, and 6, respectively. As before, estimates of span were obtained with the up-and-down method. The number of lists presented and their order of presentation were the same as in the first set of experiments. The only significant difference in procedure was that the subject was not required to arrange the response cards in their presentation order.

Since it was likely that dropping the ordering requirement would increase span, we took the precaution of increasing the size of the decks in two of the experiments. In Experiment 5, we added five cards to each deck used in Experiment 2, for a total of 15 cards/condition, and for Experiment 6, we combined the two word decks and the two picture decks used in Experi-

Table 2
Estimates of Memory Span in Experiments 4-6 (Free Recall)

Experiment	Variable	Condition	Words	Pictures
4	Phonological Similarity	Dissimilar	7.18	7.81
		Similar	5.84	6.89
5	Word (Name) Length	Short	7.60	8.67
		Lengthy	6.70	7.07
6	Irrelevant Vocalization	Silent	5.83	6.27
		Chanting	4.43	5.15

ment 3 to form a single 30-card deck for each presentation condition. Also, the presentation rate for Experiment 5 was increased to 1 card/sec, to bring it into line with Experiments 4 and 6. In all other respects, including number of subjects, the methodology for each of these experiments was identical to that for the corresponding experiment in the first set.

Results and Discussion

The results are summarized in Table 2. Given that these experiments did not require the retention of order information, it is not surprising that the observed spans were greater than those in Experiments 1-3. More important is the fact that in each experiment the speech-related variable affected word span in the expected direction and had a comparable effect on picture span.

Analyses showed that for both words and pictures the effect of the speech-related variable was reliable in each experiment [for word span, $t(12) = 3.61$ (Experiment 4), $t(15) = 3.21$ (Experiment 5), and $t(11) = 6.68$ (Experiment 6) ($p < .01$ in each case); for picture span, $t(12) = 3.71$, $p < .01$ (Experiment 4), $t(15) = 5.74$, $p < .01$ (Experiment 5), and $t(11) = 2.69$, $p < .02$ (Experiment 6)]. As before, there was no reliable evidence that the variables exerted any greater effect on word span than on picture span. In Experiments 4 and 6, the effect was larger for words, but not significantly so [$t(12) = 1.24$, $p > .10$, and $t(11) = .60$, $p > .10$, respectively]. In Experiment 5, words were somewhat less affected than were pictures.

Also replicated from the first set of experiments was the finding that, overall, spans were somewhat greater for pictures than for words. The difference was significant in all three experiments [for Experiment 4, $t(12) = 3.44$, $p < .01$; for Experiment 5, $t(15) = 3.19$, $p < .01$; and for Experiment 6, $t(11) = 2.26$, $p < .05$].

It is clear that the results of these experiments parallel closely those of the first three. It is as though, then, subjects relied on a speech-like coding of pictures, even under conditions specifically contrived to reduce the efficiency of such coding and regardless of whether they were required to remember the order in which the pictures were presented.

SUMMARY AND CONCLUSIONS

Previous research has shown that short-term memory for verbal material, whether presented auditorily or visually, is impaired when the items all sound some-

what alike, when they comprise more than a single syllable, and when the subject chants a sequence of other, irrelevant sounds during presentation of the memory items. These findings have given rise to the widely held assumption that the effective memory code for the short-term retention of verbal items is speech-like. Further, experiments reported by Conrad (1971, 1972) show that short-term memory for pictures of objects is reduced when the names of the objects are phonologically similar, which suggests that picture memory may also be mediated by a speech-like code. This suggestion gains substantial support from the experiments reported here. Within the context of immediate memory, these experiments extend Conrad's findings in three ways. First, they show that the detrimental effect that name similarity has on picture memory is comparable to the effect it has on memory for the names themselves. In other words, phonological similarity can influence picture memory just as much as it influences word memory. Second, two other speech-related variables, name length and irrelevant vocalization, were also found to affect picture memory. And again, the directions of these effects were the same as, and their magnitudes comparable to, those for word memory. Third, all of these results were obtained not only when the task required the retention of order, but also when it did not. Clearly, these results lend little support to the different-code hypothesis. To the contrary, they suggest that in immediate memory both pictures and words may be coded in the same, specifically speech-like, way.

It should be noted, however, that there is one respect in which the present results are inconsistent with a simple version of the same-code hypothesis: As in previous research (e.g., Bousfield, Esterton, & Whitmarsh, 1957; Paivio, Rogers, & Smythe, 1968; Shepard, 1967; Standing, Conezio, & Haber, 1970), the absolute level of picture memory was higher, in general, than that of word memory. Just whether and how the same-code hypothesis should be modified to account for this picture superiority effect is a debatable issue. For the present, insofar as similar findings in previous research have suggested that the short-term memory code for words is speech-like, we can conclude that the present results suggest that the code for pictures may also be speech-like.

We limit our conclusions, of course, to the materials and procedures of the present experiments. The notion of speech-like coding of pictures may apply only when the pictures are readily nameable. Moreover, even with readily nameable pictures it is unclear whether this notion is of any value for procedures involving longer term memory, or indeed for other immediate memory procedures. For instance, further research may show that it is not usefully applied to a conventional free recall procedure, in which credit is given for partial recall. And there are yet other situations (see, e.g., Brooks, 1968; Posner, Boies, Eichelman, & Taylor, 1969) in which it is certainly more useful to assume

that memory for pictures is mediated by some sort of visuospatial code. On the other hand, within the confines of an immediate reconstruction procedure in which credit is given only for completely correct performance, our conclusion in favor of speech-like coding of readily nameable pictures may well be a robust one, since each of our experiments shows evidence of such coding even under conditions specifically designed to undermine its effectiveness.

REFERENCES

- BADDELEY, A. D. Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, 1966, **18**, 362-365.
- BADDELEY, A. D., THOMSON, N., & BUCHANAN, M. Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1975, **14**, 575-589.
- BOUSFIELD, W. A., ESTERSON, J., & WHITMARSH, G. A. The effects of concomitant colored and uncolored pictorial representations of the learning of stimulus words. *Journal of Applied Psychology*, 1957, **41**, 165-168.
- BROOKS, L. R. Spatial and verbal components of the act of recall. *Canadian Journal of Psychology*, 1968, **22**, 349-368.
- CONRAD, R. Acoustic confusions and memory span for words. *Nature*, 1963, **197**, 1029-1030.
- CONRAD, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, **55**, 75-87.
- CONRAD, R. The chronology of the development of covert speech in children. *Developmental Psychology*, 1971, **5**, 398-405.
- CONRAD, R. The developmental role of vocalizing short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 521-533.
- ESTES, W. K. Phonemic coding and rehearsal in short-term memory for letter strings. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 360-372.
- LEVY, B. A. Role of articulation in auditory and visual short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 123-132.
- MURRAY, D. J. Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology*, 1968, **78**, 679-684.
- PAIVIO, A. *Imagery and verbal processes*. New York: Holt, Rinehart, & Winston, 1971.
- PAIVIO, A., & CSAPO, K. Concrete image and verbal memory codes. *Journal of Experimental Psychology*, 1969, **80**, 279-285.
- PAIVIO, A., PHILIPCHALK, R., & ROWE, E. J. Free and serial recall of pictures, sounds, and words. *Memory & Cognition*, 1975, **3**, 586-590.
- PAIVIO, A., ROGERS, T. B., & SMYTHE, P. C. Why are pictures easier to recall than words? *Psychonomic Science*, 1968, **11**, 137-138.
- POSNER, M. I., BOIES, S. J., EICHELMAN, W. H., & TAYLOR, R. L. Retention of visual and name codes of single letters. *Journal of Experimental Psychology Monograph*, 1969, **79**(1, Pt. 2).
- RICHARDSON, J. T. E., & BADDELEY, A. D. The effect of articulatory suppression in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1975, **14**, 623-629.
- ROWE, E. J., & ROGERS, T. B. Effects of concurrent auditory shadowing on free recall and recognition of pictures and words. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, **104**, 415-422.
- SHEPARD, R. N. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 156-163.
- STANDING, L., CONEZIO, J., & HABER, R. N. Perception and memory for pictures: Single-trial learning of 2560 visual stimuli. *Psychonomic Science*, 1970, **19**, 73-74.

(Received for publication November 5, 1979;
revision accepted August 1, 1980.)