Modality of presentation and blocking
in sentence recognition memory

PAUL W. FLAGG
State University of New York College, Oswego, New York 13126

and

ALLAN G. REYNOLDS
Dartmouth College, Hanover, New Hampshire 03755

In contrast to the findings of Bransford and Franks (1971) that no specific information is retained about acquisition sentences in their recognition memory paradigm, it was shown that subjects can consistently distinguish OLDs from NEWs when visual presentation is used, but not when auditory presentation is used. This is in agreement with the speculation of Katz and Gruenewald (1974) as to why such a split was observed in their study. This difference also has a bearing on the typical finding of no modality effects in long-term memory, as compared to the consistent superiority of auditory presentation in short-term memory. It was also found that blocking semantically related sentences during acquisition heightened the perceived "OLDness" of test sentences compared to nonblocked presentation.

In the middle of a conversation, if you suddenly stop speaking and ask your listener to repeat your last few sentences, after a moment of surprise, she or he will usually give an accurate summary of the sentences' "meaning." Rarely will a verbatim playback occur. This common observation (simply give it a try) is reflected in many recent studies (Barclay, 1973; Begg, 1971; Bever, 1972; Paris & Carter, 1973; Sachs, 1967, 1974). In such studies, a set of acquisition sentences is presented and then a test set is given. The subject's task is usually to discriminate the old sentences from the new. During testing, three classes of sentences are commonly presented: (1) truly old sentences which had appeared during acquisition; (2) sentences with the same or very similar meaning as some of the acquisition sentences, but a different surface structure; and (3) sentences with meanings different from any of the acquisition sentences. The typical results are a total, or near total, failure to discriminate old sentences from those similar in meaning, while sentences with different meanings are quite accurately detected.

A striking example comes from a study by Bransford and Franks (1971). As stimulus materials, they used sentences derived from a number of complex ideas, such as *The ants in the kitchen ate the sweet jelly which was on the table*. All the information in a complex idea could be described by four, simple, ONE-idea, active, declarative sentences: *The ants were in the kitchen*; *The jelly was on the table*; *The jelly was sweet*; and *The ants ate the jelly*. The simple ONE-idea sentences could be combined to yield sentences with TWO ideas, *The ants*

ate the sweet jelly, THREE ideas, *The ants in the kitchen ate the sweet jelly*, and FOUR ideas. A FOUR-idea sentence is, of course, the original complex idea. In this paradigm, ONE- through THREE-, or ONE- through FOUR-idea sentences, which were semantically related, were presented in a temporally distinct fashion during an acquisition period. During a recognition test period, OLD sentences, which had appeared during acquisition, and NEW sentences, which were novel combinations of the simple ideas, were presented. Several NONCASE sentences, which combined simple ideas across complex ideas, were presented as well.

Bransford and Franks' (1971) main finding was a total inability of subjects to discriminate OLDs from NEWs, even though NONCASEs were accurately detected. Also, as sentence complexity increased, the likelihood of an "old" response increased, regardless of a sentence's actual OLD-NEW status. This relationship is usually called the "linear effect." The explanation offered by Bransford and Franks (1971) assumed that subjects were integrating the semantically related sentences into holistic structures and storing the representation in memory. Thus, the more closely a test sentence exhausted the holistic representation, the more likely subjects were to respond "old."

The finding that subjects classified NEW sentences of greater complexity as OLD more often than they did OLD sentences of less complexity and were unable to distinguish OLDs from NEWs is robust and has been replicated a number of times (Cofer, 1973; Flagg, Potts, & Reynolds, 1975; Katz, 1973; Singer, 1973). These findings have been taken as strong evidence that subjects retain virtually nothing about the exact surface structure or physical makeup of the acquisition sentences. As a theoretical explanation, the original

integrative mechanism offered by Bransford and Franks (1971) has been accepted by most researchers in the area. Of course, some alternative accounts have been suggested; for example, Flagg (in press), Neumann (1974), and Reitman and Bower (1973) have all argued for some type of frequency model, while Katz and Gruenewald (1974) have offered a guessing strategy (however, see Flagg, in press, for evidence contradictory to this explanation).

The integrative position, of course, does not say that subjects are incapable of retaining such information; they obviously are, but the situation must be structured properly. For example, in a similar paradigm involving an incidental learning task, Graesser and Mandler (1975) found that subjects could not detect synonymous surface structure transformations in an interpretation task except when the task was to examine sentences for grammaticality. The results of Bransford and Franks (1971) are therefore taken to indicate that subjects do not *normally* retain such information.

However, one dissonant note occurs in the preceding interpretation: Katz and Gruenewald (1974) have reported a study where an OLD-NEW split occurred, with NEWs less likely to be judged as old than true OLDs of equal complexity. As Katz and Gruenewald noted, their study differed in one critical respect from the earlier studies. In their experiment, subjects read each acquisition sentence silently while the experimenter repeated it aloud; in all earlier studies, only aural presentation was used. Katz and Gruenewald suggested that the observed OLD-NEW split might be due to the difference in modality of presentation. If this is indeed the case, it suggests some interesting differences in the types of information subjects extract from visual, as compared to auditory, presentation of the same materials. Experiment 1 was a replication and extension of Katz and Gruenewald's basic findings to insure the reliability of their observed OLD-NEW split. In the present study, only visual presentation was used, in order to avoid any possible interaction that may have been present in Katz and Gruenewald's use of simultaneous visual and auditory presentation. There was a control condition in which auditory presentation was used as well.

EXPERIMENT 1

Method

Subjects. Students from introductory psychology served as subjects for course credit. First, 24 subjects were tested in the auditory condition and, then, 14 were tested in the visual condition. Testing took about 3 weeks altogether. Each subject was tested individually in a session approximately 30 min in duration.

Materials. The four complex ideas from which the stimulus sentences were derived were from Bransford and Franks (1971). For each complex idea, all possible ONEs, TWOs, THREEs, and FOURs were used, with the exception of the single THREE requiring the use of the conjunction "and." This gave a total of 12 sentences for each complex idea, 48 in all. The pool of 48

sentences was evenly divided, with respect to the number of ONEs through FOURs, into two lists. During acquisition, subjects received one of the two lists and then, during testing, all 48 sentences were presented. During testing, three NONCASE sentences, which combined simple ideas across complex ideas, and three FALSE sentences, which contained some totally new information, were presented.

Apparatus. The entire visual part of the experiment was conducted at a Beehive cathode-ray terminal connected to the Dartmouth Time-Sharing System (DTSS) via a 300 baud (30 characters/sec) phone line. The terminal keyboard was covered by a black Plexiglas response panel. Five buttons were placed in the upper left-hand corner and labeled 1 through 5; these were used in making confidence judgments. Two buttons, one labeled "old" and the other labeled "new," were placed side by side in the center of the panel (response buttons were presented only during testing). The two buttons were used to indicate whether a test sentence was identical to some acquisition sentence. Across subjects, the positions of the OLD and NEW buttons were counterbalanced.

All stimulus items were printed out of sight behind a mask covering the lower part of the terminal screen, and then advanced into view. All characters were uppercase and the contrast was adjusted so that white characters appeared on a black background. Subjects' responses to the elliptical questions asked during acquisition (see below) were recorded by a cassette tape recorder.

Procedure. *Acquisition.* Subjects were alternately assigned to one of the two acquisition lists as they arrived. Once the subject had read the instructions explaining that they would be presented a series of sentences and would be asked to answer a brief question about each, the session started. In the visual condition, each sentence remained in view for 5 sec, which is about equal to the time it takes to present a FOUR aurally, and then the screen was cleared and an elliptical question appeared (e.g., "It was? Where?"). The subject responded into a microphone on top of the terminal. In the auditory condition, sentences were read with normal intonation. Once all sentences had been presented, subjects spent 5 min crossing out all instances of a designated letter on a worksheet. No mention of a test session of any kind was made.

Test. At the start of the test session, subjects read instructions explaining that the task was to decide if each test sentence was exactly the same (word for word) as any acquisition sentence. In the visual condition, test sentences were presented one at a time and remained in view until either the OLD or the NEW button was pressed. The screen was then cleared and the word "confidence" appeared. Subjects rated their confidence in their OLD or NEW response by pressing one of the buttons labeled 1 through 5. A response of 1 indicated very low confidence and a response of 5 indicated very high confidence. In the auditory condition, subjects indicated their OLD-NEW responses and confidence ratings on a response sheet. Once a response was completed, the next sentence was read. For both test and acquisition lists in the visual condition, random orderings of sentences were program-generated for each subject so that no two consecutive sentences were from the same complex idea. In the auditory condition, four different random orderings were used for the acquisition and test sentences.

Results

As usual in this paradigm, NONCASE sentences were detected very accurately (.96 in the visual condition and 1.00 in the auditory condition), as were FALSE sentences (1.00 and .98 in the visual and auditory conditions, respectively). As the dependent measure in the following analyses, weighted confidence values were used in which the rating associated with an OLD response was multiplied by +1 and that for a NEW
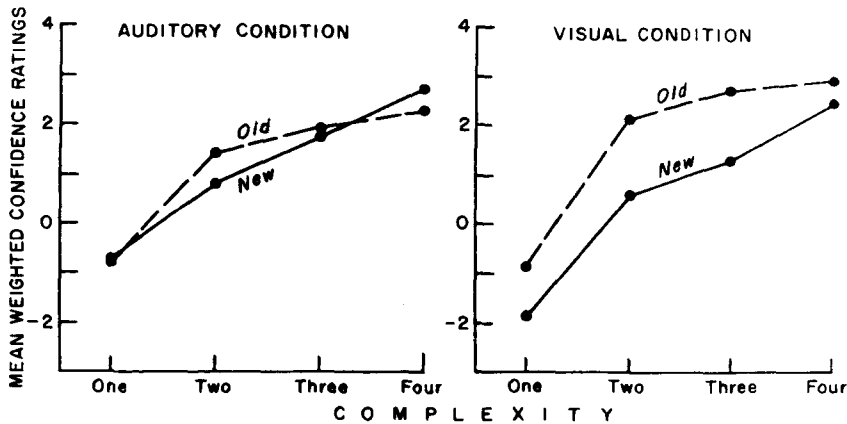
Figure 1. Weighted confidence ratings as a function of OLD-NEW status and complexity for the visual and auditory conditions of Experiment 1.

response was multiplied by −1. Analyses of variance with OLD-NEW and complexity (ONEs through FOURs) as within-subjects factors were performed on the visual and auditory data. As can be seen from Figure 1, OLDs received higher weighted confidence values than did NEWs in the visual condition [F(1,13) = 40.05, p < .001, MSe = 1.66]. As complexity increased, so did weighted confidence values [F(3,39) = 19.86, p < .001, MSe = 2.69]. The interaction was also significant [F(3,39) = 3.25, p = .03, MSe = 1.70], and, as can be seen from Figure 1, the slope for NEWs was steeper than that for OLDs. The results provide a replication of the OLD-NEW split using visual presentation, in agreement with Katz and Gruenewald (1974).

In the auditory condition, the only significant effect was for complexity [F(3,69) = 42.59, p < .001, MSe = 1.76], where weighted confidence ratings increased as complexity increased (see Figure 1). This replicates the previous results reported (Bransford & Franks, 1971).

In the present study, the acquisition period was not entirely visual, as subjects did make verbal responses to each elliptical question. To insure that an OLD-NEW split could be obtained with entirely visual presentation, a second experiment was performed.

EXPERIMENT 2

In addition to a further examination of the presentation modality, it was decided to manipulate an additional variable which might further heighten an OLD-NEW difference. This variable consisted of blocking or not blocking related acquisition sentences at presentation time. If the underlying mechanism involves integration, then blocked presentation should heighten the observed weighted confidence values. According to an integrative position, subjects are detecting the commonalities of related sentences and constructing some abstract representation (for a different account, see Flagg, in press; Neumann, 1974; Reitman & Bower, 1973). If this is the case, then blocking should make it easier to detect those interrelations, thus leading to the construction of a better

memory representation. This leads to the prediction of higher weighted confidence values for the blocked condition and possibly an interaction involving blocking and complexity, with the linear trends for levels of blocking diverging.

Method

Subjects. Fifteen subjects from the same source as in Experiment 1 were tested in each of the two levels of blocking. Each subject was tested individually in a single 30-min session.

Materials. A new set of complex ideas was constructed for Experiment 2. These complex ideas were: *The small boat fleeing the storm entered the quiet harbor*; *The colorful leaves falling from the trees covered the bare ground*; *The old train blew its whistle as it passed the small town*; and *The young mother who lost her child approached the busy policeman*. The main reason for using a new set of complex ideas was that a careful examination of the literature showed a fairly small set of complex ideas being used by a number of different experimenters. To insure generality of the findings, the additional ideas were used. A pool of 48 sentences was constructed, as in Experiment 1. For each complex idea, two ONEs, two TWOs, and two THREEs were randomly selected for inclusion on the acquisition list, for a total of 24 sentences. No FOURs were presented during acquisition. The test list consisted of three OLD sentences and four NEW sentences for each complex idea. The OLD sentences consisted of a single ONE, TWO, and THREE selected from the acquisition list. The four NEW sentences consisted of the single FOUR and a ONE, TWO, and THREE not used during acquisition. Four NONCASEs were added to the test list, for a total of 32 test sentences.

Apparatus. The apparatus used in this experiment was the same as that used in Experiment 1.

Procedure. *Acquisition*. Acquisition proceeded as in the visual condition of Experiment 1, except that elliptical questions were not asked and presentation rate was 15 sec per sentence. In the blocked condition, all the sentences pertaining to one complex idea were presented and then all the sentences pertaining to another complex idea. The orders of blocks and sentences within a block were randomly generated for each subject. In the nonblocked condition, different random orderings were program-generated for each subject, with the constraint that no two consecutive sentences could be from the same complex idea.

Test. The test session proceeded in the same manner as in the visual condition of Experiment 1.

Results

Nonblocked condition. In an analysis of variance with OLD-NEW and complexity (ONEs through THREEs)

Table 1
Weighted Confidence Values for the Nonblocked and Blocked
Conditions of Experiment 2

| | Complexity | | | | |
	1	2	3	4	Mean
	Nonblocked				
Old	1.02	1.93	1.57		1.51
New	−1.58	−.05	.97	1.61	−.22*
Mean	−.28	.94	1.27	1.61	
	Blocked				
Old	2.59	2.32	1.84		2.25
New	.28	.57	1.92	.72	.92*
Mean	1.43	1.45	1.88	.72	

*These means are based on only ONEs through THREEs and do not include FOURs.

as within-subjects factors, both main effects were significant. OLDs received higher weighted confidence values than NEWs [F(1,14) = 16.43, p = .001, MSe = 4.08; see Table 1 for the means of the nonblocked and blocked conditions]. As complexity increased, so did weighted confidence values [F(2,28) = 5.45, p = .01, MSe = 3.68]; the interaction was not significant. A separate analysis for only the NEWs gave a highly significant linear effect [F(3,48) = 8.9, p < .001, MSe = 3.27]. As can be seen from Table 1, OLD THREEs received lower recognition values than did OLD TWOs in this condition; why this occurred is not clear.

These results are in agreement with both Experiment 1 and Katz and Gruenewald (1974) and indicate that an OLD-NEW split can be consistently obtained with visual presentation.

Blocking condition. Retaining OLD-NEW and three levels of complexity (ONEs through THREEs) as repeated measures and blocking as a between-subjects factor, an analysis of variance indicated that all main effects were significant. The mean weighted confidence value for the blocked condition was 1.58, while that for the nonblocked condition was .64 [F(1,28) = 9.25, p = .005, MSe = 4.36]. This confirms the prediction that blocking would heighten weighted confidence values. The weighted confidence values for OLDs (1.87) were significantly higher than that for NEWs (.35) [F(1,28) = 26.1, p < .001, MSe = 4.02], which is consistent with the current arguments. For complexity [F(2,56) = 4.71, p = .01, MSe = 3.23], the weighted confidence values increased for ONEs through THREEs. The interaction of blocking and complexity was not significant [F(2,56) = 2.1, p = .13, MSe = 3.23]. Thus, the speculated interaction did not occur. Finally, there was an interaction between OLD-NEW and blocking-nonblocking [F(2,56) = 6.27, p = .004, MSe = 3.07]. The interaction seemed to be due to the rather unusual fact that, in the blocked condition, weighted confidence values for OLDs decreased as complexity increased. Another unusual fact was that for the NEWs in the blocked condition there was a drop from THREEs to FOURs.

DISCUSSION

As indicated by Experiment 2, blocked presentation heightened the weighted confidence values assigned to all test sentences. This may be because the manipulation made it easier to detect the commonalities existing between semantically related sentences. However, in the blocked condition, some unexpected relationships occurred. In particular, why weighted confidence values decreased as complexity increased for OLDs, rather than increasing as would be expected, is certainly unclear to the authors at this time. Whatever the reason, this result does not seem to be in line with the predictions of an integrative position (for other recent evidence against an integration position, see Small, 1975).

The most important aspect of the two present experiments was the consistent OLD-NEW split observed with visual presentation, where OLDs received higher weighted confidence values than NEWs, as compared to the absence of such a split with auditory presentation. This is congruent with Flagg (in press) and Katz and Gruenewald (1974) and is in marked contrast to the usual finding of no difference reported by all other studies in the literature using this paradigm. The only exceptions are studies by Reitman and Bower (1973) and Small (1975). However, Reitman and Bower used nonsemantic alphanumeric character strings as stimulus materials, so a direct comparison can not be made.[1] It might be noted, however, that Reitman and Bower also used visual presentation. Small used arbitrary categories of words rather than semantically related sentences. He did report OLD-NEW differences with auditory presentation, but, again, the marked differences between the types of stimulus materials used clearly prevent any direct comparisons.

The current results are entirely consistent with the speculation that the presence or absence of an OLD-NEW split when semantic materials are used in this paradigm is due to a modality effect; *visual* presentation leads to the split, while *auditory* presentation does not. It is quite possible that visual presentation makes more information available to subjects (e.g., orthographic cues) than does auditory presentation, because subjects can only store semantic information in auditory presentation. Thus, with visual presentation, subjects may be storing a holistic memory representation and some form information, while, with auditory presentation, only the former is stored. It is this additional form information which allows the discrimination between OLDs and NEWs. For OLDs, subjects qualify the information available from their holistic representation by whatever form information is available, while, for NEWs, they must rely solely on the holistic representation. This may account for the steeper slope in Figure 1 for NEWs, as compared to OLDs. For NEWs, there is no form information to qualify the powerful influence of the holistic representation. Similarly, the slope for the NEWs in Experiment 2 was quite steep compared to

that of the OLDs. In another experiment, performed by the present authors, which was identical to the visual condition of Experiment 1, except that acquisition sentences were exposed for 15 sec rather than 5 sec, exactly the same relationship held between the slopes of the OLD and NEW curves. At the same time, there was no drop between NEW THREEs and FOURs, as in the nonblocked condition of Experiment 2. From an examination of Figure 1 in Katz and Gruenewald (1974), it is evident that the relationship between the two slopes does not hold in their study. Recall, however, that they used both visual and auditory presentation; thus, direct comparison must be qualified accordingly.

The consistently observed difference between visual and auditory presentation in this paradigm is not only relevant to the issues raised by Bransford and Franks (1971) but also to the general area of research on modality effects. To quote from a recent article by Penney: "Modality differences have been reported in every short-term memory task investigated [with auditory presentation yielding superior performance] . . . Furthermore, modality effects in long-term memory have not been reliably demonstrated" (1975, p. 68). Given that the minimum time between an acquisition and test sentence was at least 5 min and could have been as much as 15-20 min, the current findings provide a long-term memory paradigm in which consistent modality effects can be obtained. The fact that such differences are observed in this paradigm may be due to the nature of the stimulus materials (semantically related sentences, rather than the more typically used word lists).

In conclusion, regardless of whether the linear relationship between complexity and weighted confidence values is due to integration or to some other mechanism, the presence or absence of an OLD-NEW split seems to be determined by the modality of presentation, at least in this paradigm. This raises a number of interesting questions about the manner in which the information present in the stimulus material is differentially coded and/or retrieved from memory. For example, what happens if the acquisition and test modalities are different? Assuming that differential coding is occurring at the time of acquisition, can the visual code be accessed when auditory testing is used and thus lead to an OLD-NEW split, or is efficient access not possible? Research is currently being conducted in an attempt to answer some of these questions.

REFERENCES

Barclay, J. R. The role of comprehension in remembering sentences. *Cognitive Psychology*, 1973, 4, 229-254.

Begg, I. Recognition memory for sentence meaning and wording. *Journal of Verbal Learning and Verbal Behavior*, 1971, 10, 176-181.

Bever, T. G. Perception, thought, and language. In J. B. Carroll & R. O. Freedle (Eds.), *Language comprehension and the acquisition of knowledge*. Washington, D.C: Winston, 1972.

Bransford, J. D., & Franks, J. J. The abstraction of linguistic ideas. *Cognitive Psychology*, 1971, 2, 331-350.

Cofer, C. N. Constructive processes in memory. *American Scientist*, 1973, 61, 537-543.

Flagg, P. W. *Semantic integration in sentence memory? Journal of Verbal Learning and Verbal Behavior, in press.*

Flagg, P. W., Potts, G. R., & Reynolds, A. G. Instructions and response strategies in recognition memory for sentences. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, 1, 592-598.

Graesser, A., & Mandler, G. Recognition memory for the meaning and surface structure of sentences. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, 104, 238-248.

Katz, S. Role of instructions in abstraction of linguistic ideas. *Journal of Experimental Psychology*, 1973, 98, 79-84.

Katz, S., & Gruenewald, P. The abstraction of linguistic ideas in "meaningless" sentences. *Memory & Cognition*, 1974, 2, 737-741.

Neumann, P. G. An attribute frequency model for the abstraction of prototypes. *Memory & Cognition*, 1974, 2, 241-248.

Paris, S. G., & Carter, A. Y. Semantic and constructive aspects of sentence memory in children. *Developmental Psychology*, 1973, 9, 109-113.

Penney, C. G. Modality effects in short-term verbal memory. *Psychological Bulletin*, 1975, 82, 68-84.

Reitman, J. S., & Bower, G. H. Storage and later recognition of exemplars of concepts. *Cognitive Psychology*, 1973, 4, 194-206.

Sachs, J. S. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, 1967, 2, 437-442.

Sachs, J. S. Memory in reading and listening to discourse. *Memory & Cognition*, 1974, 2, 95-100.

Singer, M. A replication of Bransford and Franks' (1971) "The abstraction of linguistic ideas." *Bulletin of the Psychonomic Society*, 1973, 1, 416-418.

Small, D. W. The abstraction of arbitrary categories. *Memory & Cognition*, 1975, 3, 581-585.

NOTE

1. A version of Experiment 2 was conducted using materials basically similar to those of Reitman and Bower (1973) and OLDs were found to receive higher weighted confidence values than NEWs. However, due to the many differences between semantic materials (i.e., sentences) and nonsemantic alphanumeric strings, it was felt that a direct comparison might not be appropriate.