JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR **13**, 149–162 (1974)

Verbatim and Propositional Representation of Sentences in Immediate and Long-Term Memory[1]

JOHN R. ANDERSON

Yale University

Two experiments are reported to distinguish between the representation of sentences in immediate versus longer-term memory. In Experiment I subjects studied active or passive sentences embedded in paragraph context and in Experiment II they studied isolated sentences. Their task was to immediately or later verify the truth of test sentences that were based on the study sentences. The verification latencies indicate there are two modes of sentence representation in memory. Immediately the representation is a verbatim image of the sentence but at delays there are both verbatim images and propositional representations. Evidence is offered that these propositional representations are similar to the active-like deep structure of Chomsky (1965).

What is the nature of the representation of information in long-term memory? Recently, an argument has been given (Anderson & Bower, 1973) that all information is represented in terms of abstract propositions. This contrasts with theories that assume we can remember images of sensory experience (for example, Bugelski, 1970; Paivio, 1971). Anderson and Bower deny that any long-term memories have such an imaginal character. Rather, it is argued that memories are semantic interpretations (that is, propositions) of sensory experiences. One source of support for this propositional theory is the growing evidence that we remember the gist of sentential material much better than that material's exact verbal structure (for example, Bartlett, 1932; Bransford, Barclay, & Franks, 1968; Fillenbaum, 1966; Gomulicki, 1966; Sachs, 1967; Wanner, 1968). Sachs, for instance, had subjects listen to paragraphs and try to remember whether a test sentence was the same as one they had heard 30 seconds earlier in the paragraph. Her subjects were able to detect semantic changes about 80% of the time, but could detect non-semantic changes (such as in voice of the sentence) only about 60% of the time (chance was 50%).

Sachs' experiment also indicated the need for a distinction between the immediate memory for a sentence and longer-term memory for it. She found memory for form was as good as memory for content in an immediate test (circa 90%), while memory for form but not for content drastically deteriorated with delay. Jarvella (1971) found that the ability to retrieve the exact-wording of a sentence drops off sharply after just one intervening sentence. This suggests that with no delay, there is a verbatim image of the sentence from which information about both form and content is available. The verbatim image, however, is replaced by a more permanent, more semantic representation in long-term memory.

The problem with this proposal is that it predicts no long-term memory for semantically irrelevant form and this is inconsistent with Sachs' finding of approximately 60% retention of form information. Wanner

[1] This research was supported by grant MH-13500 from the National Institutes of Mental Health to G. Bower, my adviser at Stanford University and by grant MH-24360-01 from the National Institutes of Mental Health to myself. I would like to thank G. Bower, H. Clark and L. Reder for their comments on the research and earlier manuscripts. Reprint requests should be sent to J. Anderson, Human Performance Center, 330 Packard Road, Ann Arbor, Michigan 48104.

suggested that Sachs' residual form memory could be due to the artificial task demands of the intentional learning situation which might induce the subject to deliberately propositionalize form information. For instance, the subject might state to himself and remember a proposition such as "The sentence about Galileo was in the passive voice." Thus the information about the form of the sentence is held in a structure auxiliary to that structure which contains information about the sentence's semantic content. This hypothesis will be referred to as the *Auxiliary Encoding Hypothesis*. Wanner attempted to minimize such auxiliary encodings by using an incidental learning paradigm in which subjects were not told they would be tested for memory of the material. He found perfect memory for content but no memory at all for semantically irrelevant details of the sentence. However, this result rests on few observations and so the experimental test was not very powerful.

An alternate hypothesis is that with some probability the verbatim image survives the delay, enabling exact form judgments. This *Verbatim Hypothesis* does not deny that the principal representation of the sentence in long-term memory is abstract and propositional; rather it just claims verbatim traces can exist in long-term memory. The Verbatim and Auxiliary Encoding Hypotheses cannot be distinguished in a research paradigm such as Sachs' or Wanner's in which only percent correct recognition judgments are obtained. If residual memory for form is found it will always be somewhat ambiguous whether that memory reflects deliberate encoding of auxiliary information or existence of a verbatim image or both.

It is important to find some way to discriminate between these two hypotheses. Although they are directed at explaining a weak effect in a somewhat artificial memory paradigm, a decision between the two hypotheses is crucial for evaluating the claim that all information in long-term memory is propositional. Because the physical form of a sentence is distinct from its content, a sentence memory paradigm is quite suitable for evaluating this claim.

A verification task in which reaction times are obtained may be able to distinguish between these two hypotheses about form memory. In verification tasks that test immediate memory for sentences (for example, Garrod & Trabasso 1973; Olson & Filby, 1973), it has been shown that the exact form of the to-be-remembered sentence has large effects on verification latencies. These experiments involve initially presenting the subject with sentences to remember and then asking him to verify a probe sentence based on the input sentence. It has been consistently found that probes are verified faster if the original input was of the same voice, that is, both probe and input active or probe and input passive. For example, *A girl was hit by the boy* is more quickly verified as true if the original sentence was the same than if it was *The boy hit the girl*. This is because the immediate representation of both input and probe will be form-specific and it will take longer to verify if the sentences mismatch physically.

This situation may change, however, when the information about the input sentence must be retrieved from long-term memory. Assuming that a long-term representation of the sentence is propositional, it should not reflect the voice of the input sentence. The Auxiliary Encoding Hypothesis would also assume that only this propositional representation retains the semantic content of the sentence. Information about sentence form may also be in memory, but it is recorded in a different propositional structure, separate from the propositional structure for sentence content. Therefore, according to the Auxiliary Encoding Hypothesis this information is irrelevant to the truth of the probe and the subject has no need to retrieve it in a verification task. Therefore, even though the subject may show some residual memory for the form of the sentence at delays, his verification latencies at delays should not show the interaction between input form and probe form that is present in an immediate test. There is some evidence to indicate that the interaction in verification latencies might disappear with delayed testing. Garrod and Trabasso (1973), Smith and McMahon (1970), Smith, Corriel, and McMahon (unpublished) have all found greatly diminished effects of exact form of the input sentence on verification latencies after delays of only a few seconds. However, to fully establish this finding, it is necessary to look at the interaction after greater delays such as after a few minutes of intervening activity in order to ensure that immediate memory is not being tested. As such delays, the Auxiliary Encoding Hypothesis expects no interaction at all.

In contrast to the Auxiliary Encoding Hypothesis, the Verbatim Hypothesis expects, even at delays, a significant, if reduced, interaction between voice of input and voice of probe. This is because the hypothesis claims that subjects still have available at delays a verbatim image of the sentence. Subjects can make their verification judgments on the basis of this image as well as the more abstract propositional representations. To the extent they use the former, an interaction of input and probe form will obtain. If this is the case, one would expect a significant correlation across subjects between the accuracy of exact form judgments and the degree of the obtained interaction in verification times since both phenomena rest on the availability of the verbatim image.

Experiment I

The first experiment was performed with paragraph materials so that it would be similar to the Sachs (1967) and Wanner (1968) designs which both emphasized natural comprehension of sentences in coherent text. Experiment II used isolated sentences which is typical of the sentence verification methodology (for example, Garrod & Trabasso 1973; Olson & Filby, 1972). A contrast between the results of these two experiments will be interesting as it is not obvious whether isolated sentences are represented in memory in the same way as sentences from text.

Method

The input (to-be-remembered) sentences were studied as part of short, simple, and coherently-structured stories. An example follows:

> A painter, a missionary, a cannibal, and a sailor were all on the same island together. The painter wanted to borrow the missionary's jeep to visit his friend the sailor. *The painter visited the missionary.* The painter asked to borrow the jeep. *The painter was refused by the missionary.* However, the painter stole the jeep anyway. The missionary was infuriated. *The painter was chased by the missionary.* The painter drove to see the sailor. He went into the sailor's house. He told the sailor that he had just stolen the jeep. Just then the missionray arrived. He wanted to be let in the sailor's house so he could get even with the painter. *The sailor protected the painter.* The sailor stopped the missionary at the door and refused to let him in. The missionary was not to be stopped and he had a gun. *The sailor was shot by the missionary.* The painter escaped by the back door. The sailor died in the doorway. The cannibal was a great friend of the sailor. When he heard of the murder, he searched out the painter. *The cannibal questioned the painter. The missionary was accused by the painter.* The cannibal believed the painter and swore to kill the missionary. The cannibal was a very dangerous man. *The missionary feared the cannibal.* Therefore, the missionary fled the island.

Sixteen stories of this form were constructed. Each story contained four active and four passive sentences which would be tested. The eight to-be-tested sentences in the above story are italicized. The probes testing those sentences were either actives or passives and were either true or false. False sentences were constructed by reversing the logical object and subject of the sentence. These three factors, voice of input, voice of probe, and truth were combined factorially to give eight conditions. One sentence in each story was randomly assigned to each of the eight conditions.

The 16 stories were recorded onto a tape and the subject heard them one at a time. A second tape was constructed in which the sentences that were active in the first tape were switched to passive and the passive sentences were switched to active. Half of the subjects were presented with the first tape and half with the second. Only one sequence of probe sentences was constructed. The probes were presented visually, backprojected onto a screen in front of the subject. The appearance of the probe on the screen started a timer which stopped when the subject pressed a button to indicate true or false.

The first story was used to practice the subject on the experiment. Five of the remaining 15 stories were tested under the immediate condition and ten under the delayed condition. Twice as many were tested under the delayed because more observations were necessary in that condition to achieve stable reaction times. In the immediate condition, as soon as an input sentence had occurred on the tape the tape-recorder was stopped and the probe sentence was flashed on the screen. The subject then indicated whether the probe sentence was true or false and his reaction time was recorded. If the probe was true, he was also asked whether it had the same voice as the input he had just heard. Then the tape-recorder was restarted and played until the next input sentence came up to be tested.

In the delayed condition the subject heard the entire story before being tested on the sentence. The story took about two minutes to play. The sentences were tested in the same sequence as they occurred in the story. The testing of the sentences also took about two minutes. Therefore, the lag between each input and its corresponding probe was approximately two minutes. The manner of testing of each probe was identical to the immediate condition—that is, the sentence was slide projected; the subject's verification time was taken; if the probe was true the subject was asked to decide whether its form was identical to the input.

The fifteen stories were tested in a constant order. Within each of the five blocks of three stories, one was tested under immediate and two under delayed. The conditions in which the stories were tested was counter-balanced across subjects.

In order that the subject should not come to expect a constant number of true probes, two additional sentences were randomly chosen from the story to be tested. Unlike the target sentences, these were not always transitive verb constructions. Whether both, one or none of these extra sentences were tested with true probes was randomly determined. Hence, the total number of true probes in a story would vary from four to six.

Twenty–four subjects were tested separately in a two hour experimental session. They participated as partial fulfillment of the requirement in the introductory psychology course at Stanford University.

Results

In the immediate condition subjects correctly verified 98% of the probes and, for those probes that were true, 99% were correctly identified with respect to form (that is, whether they shared the same voice with the input). Many of the 2% errors in verification were undoubtedly due to the demand for speed in the reaction time task. Subjects often volunteered the fact that they had made an error immediately after the inappropriate button-press.

In the delayed condition, subjects correctly verified 96% of the probes. Again many of the times in which they erred, they immediately corrected themselves. The reason performance on the verification task is so high at a two minute delay may be due to the coherent nature of the stories. For those probes that were true, 56% were correctly identified with respect to voice. This is significantly greater than 50% chance, $t(23) = 2.61$, $p < .01$, although not much higher.

This replicates the small residual form memory found in the Sachs' experiment. Table 1 provides a more fine-grained analysis of these voice judgments in the delayed condition. These data can be classified according to the voice of the input sentence and the voice of the probe sentence. Also the data can be classified according to whether the subject of the input sentence was in focal agreement with the paragraph. A sentence is in focal agreement if its subject agrees with the topic established by the preceding portion of the paragraph. For instance, consider the sentence, *The sailor protected the painter*, which occurs in the example paragraph. The immediately preceding context had been more about the *painter* than the *sailor*. Therefore, the sentence is not in focal agreement because its subject is *sailor*. In contrast, *The painter was protected by the sailor* would have been in focal agreement. Since half the subjects heard the active and half the passive version of each sentence, half the sentences studied were in focal agreement and half were not. Of the 120 study sentences, 32 were in focal agreement when in the passive and 88 were in focal agreement in the active.

Table 1 presents the delayed recognition data classified according to voice of input and probe, and according to focal agreement. Indicated with each proportion of correct recognition is the number of observations on which it is based. Proportion correct for form judgment is much higher in those cells where form of input and probe are the same. This indicates the strong bias of subjects to conclude that the voice of the input sentence is the same as the probe sentence with which they are being tested. The marginal proportions in Table 1 are the means of the cell proportions. These cell proportions were not weighted by their numbers of observations when obtaining the marginal means. If weighted means were used, differing numbers of observations in the main and off diagonal cells would have produced artificial differences between marginal proportions.

TABLE 1

PROPORTION CORRECT RECOGNITION OF VOICE IN THE DELAYED CONDITION OF EXPERIMENT I

		FOCAL			NON-FOCAL		
		Probe Voice			Probe Voice		
		Active	Passive	Mean	Active	Passive	Mean
Input Voice	Active	.766 ($n = 192$)	.400 ($n = 160$)	.583	.791 ($n = 48$)	.462 ($n = 80$)	.627
	Passive	.479 ($n = 48$)	.725 ($n = 80$)	.602	.349 ($n = 192$)	.643 ($n = 160$)	.496
	Mean	.622	.563	.593	.570	.553	.562

The marginal proportions reveal that recognition was much lower for sentences that were non-focal passives at input. The proportion correct for these sentences, .496, is about chance, whereas all other sentences are recognized at a mean level of .603. Comparing the difference between proportion correct for non-focal passives with all others across subjects, this difference is statistically significant, $t(23) = 2.31$, $p < .01$.[2] The effect may indicate that subjects were sometimes translating non-focal passives into actives to obtain focal agreement. The fact that recognition is not below chance indicates that subjects were not always translating passives at input. It is interesting that non-focal actives do not seem to have been translated into focal passives. In fact, the recognition accuracy for non-focal actives is slightly higher than for focal actives or focal passives. This suggests that the active is the preferred form of the sentence and subjects will not change an active sentence to achieve focal agreement.

Errors in truth judgments were excluded in computing mean verification times. They were relatively infrequent in either the immediate condition (2%) or the delayed condition (4%). The mean verification times are displayed in Figure 1, classified according to voice of input, voice of probe, truth, and delay of testing. The dots represent the actual verification times and the straight lines the predictions of a model to be presented later. An analysis of variance was performed on these data using subjects as a random variable. Each subject contributed one datum, his mean RT, to each condition in the analysis of variance. As suggested by Winer (1962, Ch. 7) a separate error estimate was obtained for each treatment based on the interaction of that treatment with subjects.

Significant main effects indicate that subjects were faster in the immediate condition than the delayed, 1.63 sec. vs. 2.63 sec., $F(1, 23) = 276.02$, $p < .001$; faster with active probes than passive, 2.04 sec. vs. 2.24 sec., $F(1, 23) = 23.23$, $p < .001$; and faster in making true responses, 2.00 sec. vs. 2.27 sec., $F(1, 23) = 63.23$, $p < .001$. They were also marginally faster when working with active

[2] This statistical test should be regarded with some suspicion as it is post-hoc. More conservative statistical procedures would not find the difference significant. However, a similar difference is obtained in Experiment II.

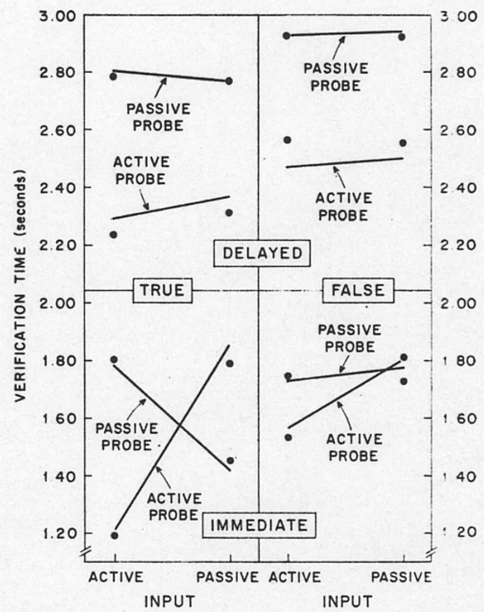

FIG. 1. Mean verification times in Experiment I as a function of voice of input, voice of probe, truth value, and delay of condition.

inputs, 2.10 sec. vs. 2.17 sec., $F(1, 23) = 3.88$, $p < .10$. There is also a highly significant lag × probe interaction, $F(1, 23) = 40.89$, $p < .001$, such that the difference between active and passive probes is much greater at long delays. There is also a significant input by probe interaction, $F(1, 23) = 14.56$, $p < .001$, such that subjects are faster when voice of probe and input match. This is the *form effect* that has been found by Garrod and Trabasso (1973) and Olson and Filby (1972). The effect is much more marked in the immediate condition as indicated by a significant delay × input × probe interaction, $F(1, 23) = 5.66$, $p < .05$. The input × probe interaction is also more marked for trues than for falses which is indicated by a significant truth × probe × input interaction $F(1, 23) = 11.72$, $p < .005$. Of course, in the case for trues, if the voice of input and probe sentences matched, the two were physically identical. When the voice matched in the false case, subject and object were reversed in the probe sentence. No other interactions were significant. Later a model will be presented that captures the variance among these data points indicated by the many significant interactions obtained.

It is tempting to conclude that there is no interaction at all between voice of the probe and voice of the input in the delayed condition. This was the conclusion offered in an earlier, cursory examination of this data (Anderson & Bower, 1973). However, there is a slight interaction, at least for the true responses. Active probes were verified 83 milliseconds faster when the input was active than when it was passive and passive probes were verified 18 milliseconds faster when the input was passive. This interaction is not at all statistically significant but it is in the direction predicted from the Verbatim Hypothesis. Given the very low memory for voice at delays (56%), it is not clear that the Verbatim Hypothesis expects a greater interaction. The Verbatim Hypothesis had expected this interaction to be directly related to the subject's ability to remember the voice of the sentence. There is only a small positive correlation, $r = .10$, between the size of the interaction and the accuracy of voice judgments for individual subjects. However, since the estimates of form memory and the verification times are quite noisy for individual subjects, a low correlation is to be expected.

Experiment II

In general, Experiment I was indecisive on the question of the Auxiliary Encoding Hypothesis versus the Verbatim Hypothesis. Because memory for form was rather poor, the Verbatim Hypothesis would not expect a very large interaction in verification times. Therefore, Experiment II was undertaken in the anticipation that it could produce better memory for form. Subjects studied sets of eight, isolated sentences rather than paragraphs embedded with eight target sentences. Memory for syntactic information is generally much higher in these types of situations (see Anderson & Bower, 1973, Ch. 8). This could be because subjects are freed from processing the theme of the story and can devote more capacity to encoding deliberate propositions about sentence form. This is what the Auxiliary Encoding Hypothesis might claim and consequently it would predict good form memory but no delayed interaction in verification times between voice of input and voice of probe. On the other hand, the Verbatim Hypothesis might attribute the better form memory to the fact that there are fewer sentences that could interfere with the verbatim image of the sentence. It would predict that improved form memory should be correlated with a greater form interaction in verification times.

Method

Except for the matter of single sentences rather than paragraphs, an attempt was made to keep the two experiments as comparable as possible. The same probe sentences were used in this experiment as in Experiment I. They were randomized into 16 sets of eight sentences such that each set had one of each of the eight possible combinations of truth value, voice of probe, and voice of input. While just one set of probe sentences was used, two sets of input sentences were constructed for study so that the same probe sentence could represent two experimental conditions depending on the input sentence. Half of the subjects studied one set of sentences and half, the other set. The input sentences were written on 3 × 5 inch cards which were studied by the subject at a 15 second rate. The probe sentences were presented as in Experiment I, that is, by a slide projector with an electronic shutter.

The first set of eight sentences was used to practice the subject on the experimental procedures. Five of the remaining sets were tested under the immediate condition in which the probe sentence was tested immediately after the input sentence. The other 10 sets were tested under the delayed condition in which a set of eight probes was tested after the subject had studied a set of eight input sentences. In the delayed condition each sentence was tested at an approximate lag of two minutes but the order of testing was not identical to the order of study. This was done to discourage subjects from trying to anticipate what the next probe sentence would be. The manner of testing was identical to Experiment I. The 16 sets were tested in a constant order. After the first practice set, one set was tested in the immediate condition and two in the delayed in each block of three. Again, the order of the conditions within each block of three was counterbalanced across subjects.

Twenty-four subjects were tested individually in a two hour session. They participated as partial fulfillment of the requirement in the introductory psychology course at Stanford. Three subjects were excluded from the analysis because the experimenter failed to execute a proper assignment of sets to the delayed and immediate conditions.

Results

The delayed voice recognition judgments for the true sentences are displayed in Table 2.

TABLE 2

PROPORTION CORRECT RECOGNITION OF VOICE IN THE DELAYED CONDITION OF EXPERIMENT II

		Voice of Probe		
		Active	Passive	Mean
Voice of Input	Active	.929	.771	.850
	Passive	.633	.781	.707
	Mean	.781	.776	.779

As predicted, there is a much higher accuracy in this experiment (78%) than in Experiment I (56%). In this table, as in Table 1 for Experiment I, there are two main results to be noted. First, subjects tended to believe that the input sentence had the same voice as the probe sentence. This is indicated by the larger values in the main diagonal cells than in the off-diagonal cells. Probably this reflects a response bias. The other noteworthy result is that subjects are more accurate in remembering the voice of active inputs than of passives. This could just reflect a bias to respond with the active. However, in Experiment I it was possible to relate the poorer memory for passives to special conditions at input. Specifically, it was found that memory was poorer for passives when they were non-focal. This input-specific effect discredits a response bias interpretation. Rather, it was suggested that subjects have a tendency to convert passives into actives at input when the passive is difficult to understand. Some subjects reported such a conversion strategy in this experiment.

Errors in truth judgments were excluded in calculating mean verification times. They were somewhat higher than in Experiment I—7% in the immediate condition and 10% in the delayed condition. This reflects, at least in part, the fact that subjects were noticeably less motivated with the unrelated materials of this experiment. The mean verification times are displayed in Figure 2, which should be compared with Figure 1 from Experiment I. Again, the dots represent the actual observations and the straight lines represent predictions of a model to be presented shortly. An analysis of variance similar to that used for Experiment I was performed on these data. The factors that proved significant in Experiment I also proved significant in this experiment. In addition to the significant effects of the previous experiment there was also a significant delay × input interaction, $F(1, 20) = 5.24$, $p < .05$, such that the active-passive difference for voice of input was greater in the immediate condition. The effect was also in this direction in Experiment I but had not been significant.

Unlike Experiment I, there was a significant effect of voice matching in the delayed condition in this experiment. Looking only at the true data, subjects were significantly faster

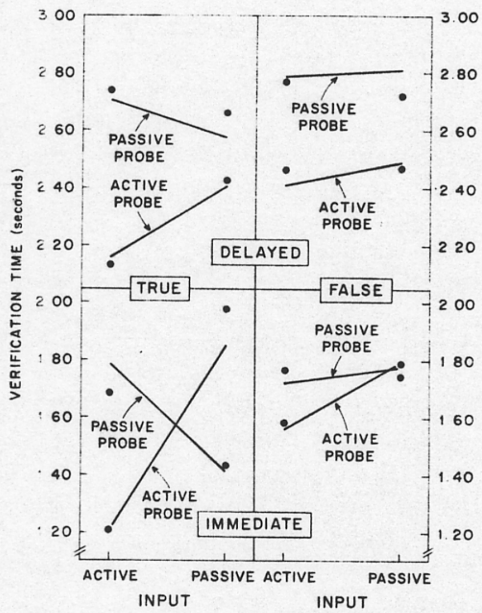

FIG. 2. Mean verification times in Experiment II as a function of voice of input, voice of probe, truth value, and delay of condition.

when voice of probe and input match, $F(1, 20) = 11.27$, $p < .005$, using the subject × delay × truth × input × probe interaction for an error term. Thus, as the Verbatim Hypothesis had predicted, an increase in the subject's ability to recognize the voice of the sentence was accompanied by an increase in the voice matching effect. There is a modest positive correlation, $r = .242$, between the accuracy of the subject's voice judgments and this interaction in verification times. If this correlation is computed over the subjects of both Experiments I and II, there is a larger range of recognition accuracies and thus the correlation is considerably larger, $r = .39$, and statistically significant, $t(43) = 3.04$, $p < .005$.

MODEL FITTING

The results of these two experiments basically reject the strong form of the Auxiliary Encoding Hypothesis which claims that there are no verbatim traces at delays. The difference between Experiments I and II can be explained by the Verbatim Hypothesis: In Experiment II there was a higher probability of a verbatim image surviving until the time of test. Let a be this probability of a verbatim image at delayed testing. It would be interesting to see whether a Verbatim Model could be fit quantitatively to the data of Experiments I and II, assuming a different value of a for the two experiments, but keeping all other parameters constant.

The Verbatim Model really specifies two distinct submodels for sentence verification, one using the verbatim image and the other using the propositional representation. Figure 3 provides a flowchart of the information processing involved in implementing the verbatim and propositional strategies within the general model. Each box in Figure 3 represents a distinct step in the processing of information and the times associated with each step are given beside the box. It is assumed that the times for each component are independent of the times for the others. In the first box the subject reads the probe sentence. This operation takes time t_1 if the sentence is active and time $t_1 + P_r$ if the sentence is passive. This extra reading time for passive sentences reflects the fact that they contain more words. After this initial step, the subject determines whether the sentence is represented in short-term memory (STM). If it is in STM he can immediately begin execution of the verbatim strategy. It is assumed that only verbatim images are retrieved for use from STM. If the subject must retrieve from long-term memory (LTM), with probability a he will retrieve a verbatim image. The time to retrieve this image is given by the parameter R. Once he has retrieved the verbatim image he will execute his verbatim strategy just as in the STM case.

The verbatim strategy was first proposed by Olson and Filby (1972) to account for their verification data and a slight variant was adopted by Garrod and Trabasso (1973). The verbatim sentence is entered into a comparison buffer where it will be compared with the probe sentence the subject is viewing. This operation of reading into the buffer takes an extra P_c milliseconds for passives, again because passives contain more words. The comparison operations can be conceptualized in template-matching terms in which words in the input are matched to words in the probe. These operations are purely syntactic and there is no reason to assume the subject accesses the semantics of any of the words he is matching. As proposed by Chase and Clark (1972) in their model for the immediate comprehension of sentences, a response index is used to keep track of the number of mismatches between input and probe. The index is initially set to true and will change every time a mismatch occurs.

It is assumed that the subject proceeds serially through the input sentence, matching it to the probe. Therefore, the first significant matching operation will concern the subjects of the two sentences. If they mismatch, the index is set to false. The time to reset the index upon a subject mismatch is M_s. The other significant matching operation concerns the verbs of the two sentences. If they mismatch, M_v milliseconds must be spent to switch the index from whatever it had been to the opposite.

There is no need for the comparison operations to proceed beyond the verb of the sentence. Since all false sentences were created by subject–object reversal, the truth of the sentence is determined at this point and the subject may execute his response index. It takes an extra F milliseconds to execute a false response. This reflects a bias in favor of positive responses which pervades most choice reaction time experiments.

To illustrate the verbatim strategy, consider the following pair of sentences, the first being the input sentence and the second, the probe:

The boy hit the girl.
The girl was hit by the boy.

The subject would first read the passive probe, then read the active input into his comparison buffer and then set his response index to true. There would be a mismatch between *boy* of the input and *girl* of the probe and the index would be set to false. The verbs would also mismatch (*hit* versus *was hit by*) and the index would be changed back to true. The subject would then execute the true index.

This verbatim strategy is described in more or less detail by many subjects as a description of what they are doing in the immediate condition. The strategy is not a general one for verification of active and passive sentences. It depends for its success on the fact that false items were created by reversing subject and object. But this is characteristic of any verification strategy that processes a sentence only at the level of a string of words. Because such strategies do not access the underlying meaning of sentence, it can only yield correct judgments about truth if there are severe constraints on the possible forms that sentences can take.[3]

To develop a model for verifying the probe against a propositional representation in LTM it is necessary to specify some properties of that representation. Anderson and Bower (1973) propose a representation similar to Chomsky's (1965) deep structure in which information is stored in LTM representations that are similar to the surface structures of active sentences. These propositional representations are abstract and both active and passive sentences must undergo some analysis to achieve the LTM format; passives however, require considerably more analysis. Similarly, the probe sentences in these experiments have to be transformed into an active propositional format before they can be used in the propositional verification strategy. Hence, it is assumed that it will take an extra P_t milliseconds to verify with a passive probe. The

[3] Note that the model, as stated, does not assume that the subject ever translates a verbatim passive into a verbatim active and stores the translation in memory. However, in Tables 1 and 2, evidence was found that subjects did sometimes translate verbatim passives. Thus, this model is only approximate in its description of the treatment of passive inputs.

reality of this extra transformation time for passives is indicated in Figures 1 and 2 by the large difference between active and passive probes in the delayed condition.

Garrod and Trabasso considered the possibility of an active LTM representation with respect to their data. They called this the Canonical representation but rejected it in favor of a Fillmore (1968) Case representation in which actives and passives are represented in a neutral format. Such a case model does not expect that there should be a greater difference between active and passive probes in the delayed condition than in the immediate. The immediate difference is a reading time effect and is reflected in the model by the parameter P_r. However, there were very strong probe voice × delay interactions in both experiments of this paper. Therefore, the Case model clearly will not do. The Garrod and Trabasso data only involved delays of a few seconds and therefore was probably not assessing a true LTM representation.

Once the probe is transformed into a propositional format it is matched to the information in LTM. Elsewhere (Anderson & Bower, 1973; Anderson, 1973) the author has described a theory of how propositional structures are matched to LTM. Details of this theory will not be discussed here except to note that it expects equal times for trues and falses in these experiments and that if the transformed probe mismatches, it can be regarded as false. In contrast, verbatim representations can mismatch and the probe still be true.

Table 3 presents the theoretical equations for verification times for the verbatim and propositional strategies. There are a number of processing times that are common to all combinations of input and probe sentences for the verbatim strategy (see Figure 3). It is not possible to obtain separate estimates of these times and therefore a single parameter represents their sum: $T_v = T_1 + t_2 + t_3 + t_4 + t_5 + t_6 + t_{10}$. Similarly, $T_p = t_1 + t_2 + t_7 + t_8 + t_9 + t_{10}$ gives the common times for the propositional strategy. The predicted verification times for the immediate condition can be obtained directly from the equations under Verbatim in Table 3. The predictions for the delayed condition are more complicated because they involve a mixture of the verbatim and propositional strategies. Let X_i be the predicted verbatim time for a particular input-probe combination and Y_i be the predicted propositional time. Then, the predicted time in the delayed condition is

$$a(X_i + R) + (1 - a)\, Y_i$$

where a is the probability of using the verbatim strategy and R is the time to retrieve the verbatim image from memory. Since a different value of a is estimated for Experiments I and II, the predictions for the delayed conditions will be different. All other parameters of the model are the same for both experiments. The parameter estimates are given in Table 4. The

TABLE 3

THEORETICAL EQUATIONS FOR VERIFICATION TIMES

Input	*Probe*	*Verbatim*	*Propositional*
The A verbed the B	The A verbed the B	T_v	T_p
The A verbed the B	The B was verbed by the A	$T_v + P_r + M_s + M_v$	$T_p + P_r + P_t$
The A verbed the B	The B verbed the A	$T_v + M_s + F$	$T_p + F$
The A verbed the B	The A was verbed by the B	$T_v + P_r + M_v + F$	$T_p + P_r + P_t + F$
The B was verbed by the A	The A verbed the B	$T_v + P_c + M_s + M_v$	T_p
The B was verbed by the A	The B was verbed by the A	$T_v + P_c + P_r$	$T_p + P_r + P_t$
The B was verbed by the A	The B verbed the A	$T_v + P_c + M_v + F$	$T_p + F$
The B was verbed by the A	The A was verbed by the B	$T_v + P_c + P_r + M_s + F$	$T_p + P_r + P_t + F$

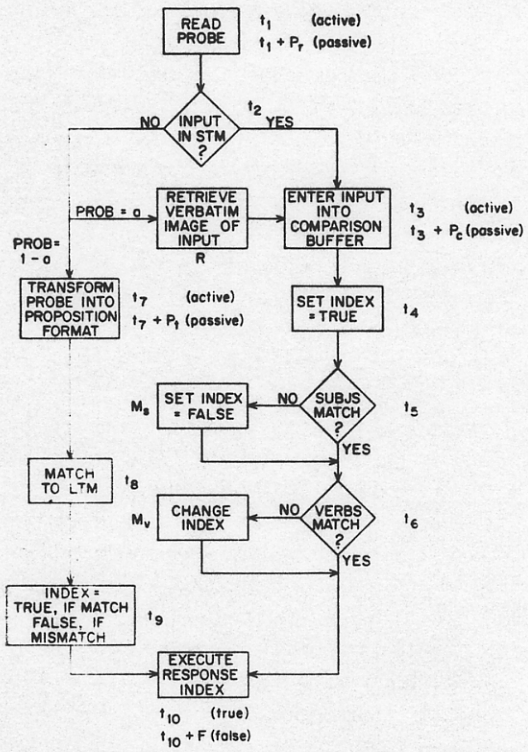

FIG. 3. A flowchart specifying the information-processing under the Propositional and Verbatim strategies.

predictions of the model are given by the straight lines in Figures 1 and 2.

The fit of the model is quite remarkable given that the same parameters are used for two experiments. The correlation between prediction and observed data is .976, indicating that 95.3% of the variance is captured by the model. With 11 parameters and 32 data points, there are 21 degrees of freedom in the predictions. Using as an error estimate the variance of individual subjects around the data points in Figures 1 and 2, the significance of the unexplained 4.7% of the variance can be evaluated. The *F*-ratio of unexplained variance to mean square error is small, $F(21, 645 = 2.78$, but highly significant, $p < .001$, because of the large numbers of degrees of freedom in numerator and denominator.[4] This significant unexplained variance arises from the attempt to generalize parameter estimates across experiments. The model yields insignificant deviations from predictions when fit separately to the two experiments. Despite the significant discrepancy, the degree of correspondence between data and theory is excellent according to current standards for verification models. Therefore, the model in Figure 3 is supported as a fairly accurate description of the information processing in these experiments.

Table 4 also gives the estimates a_1 and a_2 for Experiments I and II. These estimates do

[4] This is a very powerful test of the goodness of fit of a model to verification times. Given that an analysis of variance has been conducted on the data, the statistical test is very easy to construct. The current lamentable practice is to report how much variance is accounted for by a verification model but to give no indication of the significance of the unexplained variance.

TABLE 4

PARAMETER ESTIMATES

Common Verbatim Times	T_v	1207 msec.
Added Reading Time for Passives	P_r	70 msec.
Added Comparison Time for Passives	P_c	141 msec.
Subject Mismatch Time	M_s	205 msec.
Verb Mismatch Time	M_v	299 msec.
Time to Retrieve Verbatim from LTM	R	649 msec.
Common Propositional Times	T_p	2350 msec.
Added Transformation Time for Passives	P_t	427 msec.
Added Response Times for Falses	F	155 msec.
Probability of Verbatim Strategy in Experiment I	a_1	.106
Probability of Verbatim Strategy in Experiment II	a_2	.348

not give the probability that a verbatim image is available to the subject, but rather the probability that a verbatim image is *both* available and used to verify the probe. It is conceivable that the subject will use a propositional representation to verify the probe but still have the verbatim image available to make subsequent judgments about sentence form. Thus the a estimates provide a lower bound on the availability of a verbatim image for a form judgment. If a were the probability of a verbatim image, then the probability of a correct form judgment would be $a + 1/2(1 - a)$, using the conventional guessing model. This would predict 55.3% correct form judgments in Experiment I and 67.4% in Experiment II. The observed percentages were 55.8% and 77.9%, respectively. So the a_2 estimate considerably underpredicts voice accuracy in Experiment II. This may either indicate a_2 is an underestimate of the availability of a verbatim trace or that subjects sometimes make form judgments on the basis of information other than a verbatim trace. Perhaps they were also recalling propositions about sentence form in a manner proposed by the Auxiliary Encoding Hypothesis. Some subjects do report using such auxiliary information. It is possible then, that a subject's form judgments are based upon these auxiliary encodings as well as verbatim traces. Regardless of whether or not auxiliary encodings are also used, the conclusion of these experiments is that the subject does have available nonpropositional, verbatim traces.

Conclusions

This research indicates the existence of two modes of information representation in memory. Evidence was found for a verbatim image of the sentence which survives for at least several minutes with small but not insignificant probabilities of being employed. The second representation is an abstract propositional one that seems much more durable. Its greater durability is indicated by the high accuracy in truth judgments at delays. The propositional representation also seems to have an active-like character as indicated by the slow verifications to passive probes at delays. It is tempting to think of the verbatim image as a relatively unanalyzed acoustic (phonetic) or visual (orthographic) image of the sentence. The problem with this assumption is that there is evidence for syntactic organization in the immediate memory of sentences. Walker, Gough, and Wall (1968) found that it is easier to retrieve information within a phrase than across phrase boundaries of a sentence. Also, sentence span is longer than unrelated word span (Miller, 1956). If we assume that the subjects are using a verbatim image in such immediate memory tasks, then these experiments provide evidence that the image has some syntactic structure. However, an alternate assumption would be that immediately, as well as at delays, there is both a propositional and a perceptual representation of the sentence. Immediately, the perceptual is highly available producing good memory for exact wording. Nonetheless, the propositional is sufficiently available to produce the propositional effects. At delays, however, the propositional representation comes to dominate and effects of sentence form diminish. This hypothesis comes close to obliterating the short-term versus long-term memory distinction in favor of a model in which there are faster decay rates for the perceptual than for the propositional information. This point of view has been advanced by a number of theorists (Craik & Lockhart, 1972; Melton, 1963; Murdock, 1972; Tulving, 1968).

A final remark is needed concerning the difference between Experiments I and II. One hypothesis might have been that the paragraph material in Experiment I would be processed in a different way than the single sentence material of Experiment II (see Craik & Lockhart, 1972). However, it does not seem to be the case that sentences are encoded at a fundamentally different level when embedded in paragraphs. It seems clear that there are two types or levels of encoding for a sentence, but they appear to be used in both experiments. All that seems to have changed between the two experiments is the proportion of verbatim images available in the delayed condition. Their lesser availability in Experiment I can perhaps be attributed to the greater processing capacity demands in forming semantic connections among sentences and the additional interference to verbatim images caused by the extra sentences.

References

Anderson, J. R. Retrieval of propositional information from long-term memory. *Cognitive Psychology*, 1974, **5**, in press.

Anderson, J. R., & Bower, G. H. *Human associative memory*. Washington: Winston and Sons, 1973.

Bransford, J. D., Barclay, J. R., & Franks, J. J. Sentence memory: A constructive versus interpretative approach. *Cognitive Psychology*, 1972, **3**, 193–209.

Bartlett, F. C. *Remembering: A study in experimental and social psychology*. Cambridge: University Press, 1932.

Bugelski, B. R. Words and things and images. *American Psychologist*, 1970, **25**, 1002–1012.

Chase, W. C., & Clark, H. H. Mental operations in the comparison of sentences and pictures. In L. Gregg (Ed.), *Cognition in learning and memory*. New York: Wiley, 1972.

Chomsky, N. *Aspects of the theory of syntax*. Cambridge, Mass.: MIT Press, 1965.

Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 671–684.

Fillenbaum, S. Memory for gist: Some relevant variables. *Language and Speech*. 1966, **9**, 217–227.

Fillmore, C. J. The case for case. In E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory*. New York: Holt, Rinehart and Winston, 1968.

Garrod, S., & Trabasso, T. A dual-memory information processing interpretation of sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 155–167.

Gomulicki, B. R. Recall as an abstractive process. *Acta Psychologica*, 1956, **12**, 77–94.

Jarvella, R. J. Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 409–416.

Melton, A. W. Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 1–21.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, **63**, 81–97.

Murdock, B. B., Jr. Short-term memory. In G. H. Bower (Ed.), *Psychology of learning and motivation*, Vol. 5. New York: Academic Press, 1972, pp. 67–127.

Olson, D. R., & Filby, N. On the comprehension of active and passive sentences. *Cognitive Psychology*, 1972, **3**, 361–381.

Paivio, A. *Imagery and verbal processes*. New York: Holt, Rinehart, and Winston, 1971.

Sachs, J. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and psychophysics*, 1967, **2**, 437–442.

Smith, K., Coriell, J., & McMahon, L. E. The role of active and passive voice in answering questions. Unpublished manuscript, Bowling Green State University, 1971.

Smith, K., & McMahon, L. E. Understanding order information in sentences: Some recent work at Bell Laboratories. In W. J. M. Leveit & G. Flores d'Arcais (Eds.), *Advances in psycholinguistics*. Amsterdam: North-Holland, 1970.

Tulving, E. Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

Walker, E., Gough, P., & Wall, R. Grammatical relations and the search of sentences in immediate memory. Paper presented at the meetings of the Midwestern Psychological Association, 1968.

Wanner, H. E. On remembering, forgetting, and understanding sentences: A study of the deep structure hypothesis. Unpublished doctoral dissertation, Harvard University, 1968.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received July 13, 1973)