Forgetting of verbatim information in discourse

GREGORY L. MURPHY
University of Illinois, Urbana, Illinois

and

AMY M. SHAPIRO University of California, Berkeley, California

In three experiments, the causes of forgetting of verbatim information in discourse were investigated. In Experiment 1, we tested the view that the surface form of a sentence decays when linguistic elements are integrated into larger discourse structures. Contrary to such a view, the results showed that when text could be integrated, both content and verbatim memory improved. In Experiment 2, we examined the possibility that task instructions could influence the level of memory for exact wording and content. The results showed that although task did influence the amount of verbatim memory, it did not affect memory for content. Experiment 3 was an investigation of the degree to which subjects would spontaneously encode surface information on the basis of the social interactiveness of the sentences. Past research has shown that verbatim memory for insults, jokes, and other personal utterances is quite good. In this experiment, identical sentences were tested in texts that had high versus low interactiveness. Verbatim memory was much higher for the same sentences in the highly interactive context. A pragmatic account of verbatim memory is given to explain how content and instructions influence the encoding of surface form.

Since the seminal work of Sachs (1967), it has become a commonplace that memory for the surface form of text (verbatim memory) is rather poor. Sachs showed that readers could recognize the exact wording of a sentence only with immediate testing; after 80 syllables of intervening text, their verbatim memory had faded, even though their memory of the meaning of the sentence was virtually intact. This demonstration has been repeated numerous times with variations (J. R. Anderson, 1974; Bransford, Barclay, & Franks, 1972; Gernsbacher, 1985; Graesser & Mandler, 1975; Jarvella, 1971, 1979; Johnson-Laird & Stevenson, 1970; Sachs, 1974, among many others). Although this phenomenon is very familiar, explanations for the loss of verbatim memory are not as well established. Authors agree that verbatim information is not as important as the content of a text, and so does not need to be retained in most cases, but this is not an explanation of how it is forgotten. In the three experiments reported here, we evaluate possible explanations of the fragility of verbatim memory within written text or spoken conversation.

Why study verbatim memory? Memory for surface features can be important in and of itself, as Kolers (1975) noted:

This research was supported by NIMH Grant MH-41704. The authors are grateful to Barbara Skalak, whose earlier experiments on this topic were very helpful in designing the present study. We thank Frank Bellezza and anonymous reviewers for helpful comments on a draft of this article. Correspondence may be addressed to G. L. Murphy, Department of Psychology, University of Illinois, 603 East Daniel St., Champaign, IL 61820.

A particular brush stroke, curvature of a line, or seam of a casting; reflectance, shade or crackling of a varnish; choice of a word or its syntactic embedding—let the expert tell one work from another, one make from another, often at a glance [Surface features'] perception can be of the utmost significance to interpretation. Indeed, despite their superficial nature they can even be the message. (p. 699)

We will discuss the last point at greater length later. One of the ways that surface information can be important is in language processing. Although the main function of language is to communicate ideas—not surface forms—the retention of surface forms can be a crucial step in formulating and understanding utterances (as pointed out by Jarvella, 1979; see also Levelt & Kelter, 1982). For example, it has been argued that understanding some anaphors requires listeners to retain the exact wording of their antecedents (Murphy, 1985b; Sag & Hankamer, 1984). Furthermore, making an antecedent's surface form less accessible does slow down processing of the anaphor (Levelt & Kelter, 1982; Murphy, 1985a). Also, focusing techniques in discourse may have the effect of causing better surface memory for the focused elements (Birch, 1993; Malt, 1985). Thus, short-term surface memory may be important to the on-line interpretation of language.

Another reason for studying verbatim memory concerns not only its role in on-line processing, but also its part in global comprehension and long-term memory. Any theory of memory for discourse should account for the aspects of discourse that are retained and should provide an account of their representation. Thus, it is necessary to have an explanation of why verbatim information is usually poorly retained. And it is even more important to have an explanation of cases in which it is accurately remembered, because they are likely to be crucial test cases of our theory of memory for language. Finally, there are important applications of people's ability (or lack thereof) to remember verbal material verbatim. For example, eyewitness testimony may depend on excellent memory for a conversation, following instructions may require remembering the exact wording, and social interactions may depend on the precise surface form that was used in making a socially sensitive utterance.

Given our interest in how memory for surface information reflects actual language use, we have derived three desiderata, or criteria, that will help determine whether a given experiment is relevant to verbatim memory of normal discourse. (1) There must be separate measures of verbatim and content memory. Almost all verbal learning studies that require subjects to recall word lists could be considered "verbatim memory" studies, because they require the exact words to be produced. But failure to recall a word in such a situation could reflect the loss of verbatim information alone (as when a synonym is recalled), loss of content, or both. Finding that a variable influences subjects' ability to remember exact wording without any comparable measure of content does not show whether the variable affects performance as a whole, or memory for surface form in particular. (2) The stimuli must form a realistic, connected text. Individual words or sentences are unlikely to be processed in the same way as actual discourse, as years of research (since Bartlett, 1932) in discourse memory have shown. Thus, results from memory for word lists simply may not extend to memory for text or conversation. (3) Subjects must not be told to memorize the text. More generally, the acquisition phase must be as similar to normal comprehension as possible. Memorization strategies are not the same as comprehension processes, so when subjects know that they are to receive a memory test (especially a verbatim memory test), the results may be quite different from normal comprehension (as shown by Johnson-Laird & Stevenson, 1970; Levelt & Kelter, 1982, Experiment 5).

We should emphasize that we are not criticizing previous studies that did not follow these three criteria. In most cases, they were designed to answer questions that are different from the ones we are posing here. Our argument is simply that, to understand verbatim memory in normal language processing, these criteria seem desirable, and so studies that violate one or more of the criteria will be somewhat suspect in what they tell us about memory for discourse. Of course, it is an empirical question as to whether studies that don't follow these criteria yield results that are different from those that do.

Explanations of Poor Verbatim Memory

The finding of poor verbatim memory is so widespread that textbooks often describe it as a basic fact about memory, without providing any explanation of the finding. One could argue that early levels-of-processing theory (Craik & Lockhart, 1972) took poor memory for surface form as a primitive, unexplained element in the theory, rather than attempting to account for it. Gernsbacher (1985) provides a very detailed and useful description of explanations of the loss of verbatim memory. We present two of the accounts she describes, which will be tested in the experiments.

The view Gernsbacher (1985) calls the integration hypothesis proposes that as linguistic elements appear, they are encoded into larger structures. Words are encoded into clause and sentence meanings; sentences are encoded into larger discourse units or mental models. This encoding process creates more efficient memory structures, which are placed into long-term memory. The smaller constituents, out of which these structures were built, then decay. Evidence for this view is described in Experiment 1. The most familiar support for it is the work of Bransford and Franks (1971; Bransford et al., 1972), which showed that subjects remembered the situation described by a set of sentences, but not the individual sentences (see also Mani & Johnson-Laird, 1982, p. 185). In short, according to the integration view, the consolidation of smaller units into larger units results in the loss of the smaller units. (Note that the integration view is an explanation of the loss of verbatim information. Many other researchers have studied the beneficial effects of integration on comprehension of memory without making any claims about its influence on verbatim memory. In the present study, we address the more specific hypothesis about forgetting surface form.)

Gernsbacher (1985) suggested a new explanation for verbatim memory loss, called the processing shift hypothesis. According to this view, the memorability of any item is a function of the attention that is focused on that item. As long as a subject is attending to a particular structure in memory, his/her memory for the parts of that structure will be excellent. When a new structure is built, as when a new episode in a story is encountered, the previous structure begins to decay. Thus, loss of verbatim memory results from leaving a given substructure in memory and allocating attention to a new one. On the face of it, this would seem to entail a loss of content memory as well. However, Gernsbacher argues that surface form is the fastest changing kind of information in most text, and is therefore subject to the most processing shifts. Because processing takes place at a number of different levels (e.g., word, phrase, sentence, event), there can be a hierarchy of processing structures that influence memory. Because surface information is at the bottom of the hierarchy, memory for it is poorest. The main strength of this view is that it can explain why surface memory loss seems to occur immediately after structural boundaries in sentences and stories (Clark & Sengul, 1979; Gernsbacher, 1985; Jarvella, 1971, 1979).

Both of these explanations have considerable difficulty accounting for an important phenomenon that was first documented by Keenan, MacWhinney, and Mayhew

(1977). They discovered that the audience in a seminar could remember the verbatim wording of some of the sentences produced in the seminar when tested 27 to 48 h later. The subjects did not remember verbatim the sentences that contained pallid information or factual statements. Rather, they remembered the exact wording of jokes, personal statements, insults, and the like. This effect was replicated by Kintsch and Bates (1977) and Mac-Whinney, Keenan, and Reinke (1982) for even longer intervals. In Keenan et al.'s terms, subjects remember high interactional content sentences, or sentences that focus on "information about the speaker's intentions, his beliefs, and his relations with the listener" (p. 550). (We will refer to these sentences as highly interactive sentences.) Keenan et al. suggest that remembering the exact wording of these sentences could be important to one's future interactions with and social evaluations of speakers. A series of control studies revealed that the highly interactive sentences themselves were not more memorable when presented in isolation, but that being part of a conversational interaction was crucial. Other studies have confirmed that surface forms of highly interactive text are remembered better than low interactive text (Gibbs, 1981; Gibbs & Nagaoka, 1985; Murphy, 1992).

Although this kind of effect seems to have very significant implications for our understanding of text memory, its influence has been weakened by the absence of a theoretical framework to explain it. Keenan et al. (1977) give a number of functional explanations for why highly interactive sentences should be remembered better, but they do not provide a mechanism that explains why verbatim memory is normally so poor, nor do they spell out the encoding or retrieval mechanisms that lead to better memory for highly interactive sentences. By the same token, the integration and processing shift hypotheses described earlier do not explain why surface memory is better for these sentences.

To address this lack of a processing explanation, we propose an account that we call the pragmatic view. In brief, the pragmatic view proposes that listeners attend to the level of analysis of text that is most relevant, important, or salient, given their current goals. This attentional allocation greatly affects which information will be recalled later. Thus, this view is similar to the processing shift theory in that it relies on the allocation of attention to explain memory differences. Unlike that view, it does not attribute those differences to structural variables, but rather to strategic control by the listener and the particular features of the text. In most cases, listeners are interested in the content of a conversation, and that is where they allocate their attention. However, when the exact wording of a sentence has crucial implications for the goal at hand, it may receive as much attention as the meaning of the sentence, and as a result, it may be well remembered. Furthermore, certain sentences attract attention to their surface forms, especially those that include social information of some kind. One's reaction to a compliment depends on the exact adjective that was used; one's outrage to an insult depends on the precise epithet that was used and the way it was uttered. When such utterances are heard, people attend to and remember their exact wording. The pragmatic view also acknowledges that goals at retrieval can influence the amount and kind of information that is retrieved (Morris, Bransford, & Franks, 1977), though encoding obviously provides a bottleneck that can limit the effectiveness of any retrieval strategy. Also, it should be emphasized that the pragmatic view does not deny that structural memory effects (e.g., interference from similar materials) influence surface memory. However, this view suggests that those effects are qualitatively similar in memory for surface form and memory for content (see J. R. Anderson & Paulson, 1977).

The main contribution of the pragmatic view is to link together two different ideas in order to explain forgetting of surface form. One idea is the general notion that the information that is attended to at encoding is the information that will be better remembered. The other idea is that there is something special about highly interactive sentences. Our claim is that such sentences can cause readers or listeners to pay more attention to their exact wording than they would for normal sentences. Unfortunately, we cannot specify what makes a sentence highly interactive much more specifically than Keenan et al. (1977) did. Clearly, utterances with social import are examples, because their exact wording is crucial to their interpretation. However, it is also likely that unusual or unconventional language will attract attention to itself (Gibbs, 1981). To a large degree, whether a sentence is highly interactive is an empirical question that needs to be answered by psycholinguistic research rather than by stipulation.

In the experiments presented here, we examine the three hypotheses just described for why verbatim memory is poor. Although these views have been presented as competitors, it is by no means clear that they are mutually exclusive. Each one might account for some part of the loss of verbatim information. The present experiments focus on predictions made by the integration and pragmatic views, but the question of whether other factors are necessary to account for all the extant data is presented in the General Discussion.

EXPERIMENT 1

One of the most popular explanations for poor verbatim memory is the integration hypothesis (Gernsbacher, 1985). The work of Bransford, Franks, and colleagues is often interpreted as supporting this view. For example, Bransford and Franks (1971) presented subjects with sentences, some of which described a single situation, such as: "The jelly was on the table"; "The ants ate the sweet jelly which was on the table"; and "The ants in the kitchen ate the jelly." Their striking finding was that the subjects could not identify which sentences from this set had been

presented. Instead, they "recognized" each sentence in proportion to the number of propositions from the original stimuli it contained. Bransford and Franks concluded that the subjects did not remember the individual sentences, but instead had combined them into a more complex memory structure. As a result of this integration, information specific to individual items (i.e., verbatim information) was largely lost. Bower, Black, and Turner (1979) and Bransford et al. (1972) explained their results in similar ways.

An important point about the integration hypothesis is that it contradicts the notion that memory should be better when material is more integrated (e.g., Bransford & Johnson, 1973; Craik & Tulving, 1975). It is not clear why integration should have a deleterious effect on surface form when it has a beneficial effect on content memory. There is some evidence against this claim from a nonexperimental source: when actors learn their lines for a part in a play, they focus on integration and elaboration of the meaning of the lines, rather than use rote memorization (Noice, 1992). That is, integration appears to aid their verbatim memory.

An empirical problem is that many tests of the effect of integration have not clearly separated verbatim and content memory, and so do not provide clear evidence for or against this explanation. For example, in Bransford and Franks's (1971) experiment, the recognition foils had meanings and wording that were both different from the old sentences. Thus, errors represent a failure to remember both the meaning and surface form of the sentence. The same objection holds for a study by Peterson and McIntyre (1973), who specifically tested the integration hypothesis (see also Pompi & Lachman, 1967). The recognition task used in the present study provided a measure of verbatim memory per se, as well as giving an estimate of memory for content.

Anderson and Bower (1973, chap. 8) describe an experiment that seems to give strong evidence for the integration theory. Their subjects studied sentences and then were asked to verify related statements, immediately or after a delay. The statement in the verification task was either identical to the originally studied sentence, or it reversed the voice (active or passive) of the sentence. Anderson and Bower's results in the delay condition showed that when the studied sentences formed a story, there was little effect of surface form of the test sentence on verification. But when the studied sentences were unrelated. there was a reliable matching effect: The subjects responded more quickly when the test sentence matched the form of the studied sentence. This result is just as the integration theory would predict. Unfortunately, the study is not described in much detail, and certain crucial details are not explicitly described. For example, there was a reliable matching effect in the isolated-sentence condition, but there was also evidence of a small effect in the story condition, and no statistical test was reported to show that the two patterns differed reliably. In addition, the nature of the task seems to have encouraged the subjects

to memorize the sentences, because they were asked to identify the voice of the original studied sentences during the verification task (p. 226; this was also the technique used by J. R. Anderson & Paulson, 1977). Thus, it is not clear that the results can be extended to apply to more normal comprehension.

More recently, Gernsbacher (1985, Experiment 5) tested the integration hypothesis by using pictures that created a story. She presented the pictures in either their normal or a scrambled order and found, contrary to the integration hypothesis, 6% better memory for surface information in the normal order. The test of surface information was whether a picture was presented in its normal or a mirror-image orientation. Her Experiment 6 replicated the result with verbal materials replacing the pictures (constituent order was used as surface information).

Gernsbacher's results, then, provide evidence against the integration hypothesis. Nonetheless, we wanted to perform a similar experiment under slightly different conditions to meet the criteria we outlined in the introduction. In Gernsbacher's methodology, subjects were shown a picture and its mirror image before the start of the experiment proper and were told that they would have to say which one was presented. (An analogous instruction was used for the verbal materials.) Also, the subjects were told that they would have to summarize the stories. Thus, her experiment does not fulfill all the criteria listed earlier for our investigation of verbatim memory; the subjects knew that they were participating in a memory test, and that their verbatim memory in particular would be tested. Furthermore, Gernsbacher did not have an equivalent measure of content memory. (Although she asked the subjects to summarize stories, the scores on this test are not comparable to accuracy of recognition in the verbatim test.) Again, we emphasize that these are not necessarily shortcomings in experimental design, but that they simply do not meet the goals of the present investigation. Thus, Experiment 1 provides a test of the integration hypothesis under the conditions outlined previously.

Method

Subjects. The subjects were 24 Brown University undergraduates who participated for payment.

Materials. Two stories were written for the experiment. One, entitled "First Day of Practice Teaching," was 23 sentences long, describing the experiences of a student teacher who taught his first class in elementary school. The other was entitled "The Playoff Game," and it described the experiences of a character attending his daughter's softball game. It was 24 sentences long. Each story was written in two versions that differed only in the surface forms of seven of the sentences. For example, one version contained the sentence "James introduced himself and then took attendance," whereas the other contained "After introducing himself, James took attendance." To the degree possible, the alternate versions of these sentences contained the same words rearranged into different (but semantically equivalent) syntactic structures.

The stories were recorded onto audiotape for presentation. In the "normal" condition, the playoff game story was presented first, followed by the practice-teaching story. Each story was preceded by its title. In the "jumbled" condition, the sentences of the two

stories were intermixed and randomized. We expected that intermixing the stories would prohibit integration more than would reordering alone. There was a short gap between the two halves of the presentation, but no title was given for either half. In total, there were four versions of the materials recorded: Versions A versus B, crossed with the normal versus jumbled conditions. The presentation times of all versions were equal.

A test of 21 items was also constructed. It consisted of 7 "old" items, 7 semantic foils, and 7 surface foils. The old items had actually occurred in the story exactly as presented. The semantic foils did not occur in the story, and in fact contained parts of sentences from the two different stories. For example, the item "Brian tried to get there early so he'd be able to get a seat, but he arrived to find the parking lot already crowded," contained a reference to a crowded parking lot from the practice-teaching story and a reference to trying to get there early from the playoff game story. Thus, the subjects could not identify semantic foils simply by identifying new material that had not occurred in either story; they had to identify whether that particular combination of ideas had occurred in a sentence. (Previous experiments showed that when easier semantic foils were used, subjects in all groups performed near ceiling.) The surface foils were the alternate versions of a sentence that occurred in one of the stories. Because of the counterbalancing of the two versions, a given test item was a surface foil for half of the subjects and an old item for the other half. Thus, the naturalness of or preference for a given sentence could not influence the measure of surface memory across the two versions (see Bock & Brewer, 1974).

Procedure. The subjects were told that they were participating in an experiment on language comprehension. They were told to listen to two stories that were to be played on a cassette recorder. The only instructions they were given were, "Please pay close attention to what you are about to hear, as the passages may be hard to understand. Later, you will be asked some questions about how well you understood them, how much you liked them, and so forth.' Following this, one of the recordings was played for the subjects. They were then given written instructions for the test, which emphasized that they were to identify sentences on the test that were absolutely identical to those they had heard in the story. No time limit was imposed. Underneath each test item was a scale of 1-7, which the subjects used to indicate their confidence in their answer. Because the results were quite clear from the recognition accuracy data, we did not analyze the confidence judgments. Reading the test instructions and answering the experimenter's questions took about 3-5 min, which was deemed enough time to clear the last part of the text from short-term memory (Sachs, 1967).

Results and Discussion

The subjects' responses were divided into content and surface memory scores. The content memory score was simply the percent accuracy on the semantic foils. We expected this score to be quite high for both groups. The surface memory score was the number of hits (responding "yes" to the old items) plus the number of correct rejections of the surface foils. According to the integration theory, integration should help content memory but should hurt verbatim memory, because the integrated memory structures do not preserve the original sentence structure.

As Table 1 shows, the subjects in the normal condition scored 20% higher on the content memory test items [t(22) = 3.10, p < .01]. Perhaps not surprisingly, when the stories were presented in a comprehensible manner, the subjects were able to encode and remember the mean-

Table 1
Percent-Correct Memory for Surface and
Semantic Information in Experiment 1

	Story Condition	
	Normal	Jumbled
Memory for meaning	93	72
Surface memory	62	52

ing better. This result is important, however, because it shows that the stimulus manipulation was strong enough to influence memory. According to the integration view, this advantage should be reversed for verbatim memory. Instead, verbatim memory in the normal condition was about 10% higher. This difference was only marginally reliable [t(22) = 2.04, p < .10], but the main point here is that the difference is in the direction opposite to that predicted by the integration view. A similar experiment in our laboratory also showed a slight advantage in verbatim memory when the material was more easily integrated. 1

The results of this study contradict the prediction of the integration view. As the material became more easily integratable, surface memory did not decay. These results are consistent with those of Gernsbacher (1985). However, unlike her study, our subjects were not aware that memory would be tested.

These results may appear to call into question the findings of Bransford et al. (1972) or Bransford and Franks (1971), who seemed to find poor verbatim memory with integratable material. However, as mentioned earlier, these studies did not provide separate estimates of verbatim and content memory. For example, Bransford and Franks showed that subjects false-alarmed to complex sentences such as "The ants in the kitchen ate the sweet jelly which was on the table," which had never been presented. However, this sentence is both semantically and syntactically distinct from any sentence presented in the memory set; it is not a test of surface memory alone. Bransford et al.'s results were of a similar form. We pointed out previously that organization or integration of materials is generally believed to improve memory. Thus, the integration view is, in some respects, counterintuitive. The present results and those of Gernsbacher (1985) cast doubt on its claim: If anything, integration has a slightly positive effect on verbatim memory.

EXPERIMENT 2

Experiment 1 provided evidence against the integration view; we next wanted to investigate predictions of the pragmatic view. As described earlier, the essence of the pragmatic view is that different orienting tasks or certain characteristics of a text can cause listeners or readers to focus on the surface aspects of a discourse. If they do so, their surface memory will improve.

Graesser and Mandler (1975) performed an experiment that provides considerable support for this view. In their Experiment 2, subjects judged either the grammaticality or interpretability of sentences in a passage. (Ungrammatical and nonsensical sentences were inserted to provide negative responses.) A forced-choice memory test later showed that the subjects who judged grammaticality were reliably less likely to choose distractors that differed only in surface form. Verbatim memory was above chance for the grammaticality group, but not for the interpretability group. This study is very close to fitting the criteria described in the introduction; the main exception is that the subjects evaluated every sentence separately rather than simply understanding the passage or evaluating it as a whole. Also, the filler sentences decreased the realism of the passage by the addition of nonsensical sentences. Nonetheless, this study provides evidence for the pragmatic view.

Experiment 2 had two main goals. The first was to replicate the major result of Graesser and Mandler (1975), which has not received sufficient attention in the literature. We used a slightly more naturalistic task and set of materials. The second goal was to attempt to influence both content and surface memory through orienting tasks. Just as verbatim memory might be improved by encouraging subjects to encode surface information more actively, memory for content might be improved or degraded by appropriate instructions. Graesser and Mandler found no effects on memory for meaning, but in the present experiment we used different orienting instructions and a more detailed test for content. That is, it explicitly tested subjects' memory for the ideas mentioned in the text rather than using errors in the verbatim test as the content memory measure.

Our expectation, then, was that instructions could influence both surface and content memory. Furthermore, we once again did this within the confines of the criteria described in the introduction. That is, the subjects were not told that they were involved in a memory experiment, and they did not evaluate each sentence separately.

Method

Subjects. Thirty-two Brown University students served as paid subjects.

Materials. Two basic narratives were written for the experiment. Each was purportedly an essay by a high-school student comparing the careers of Charles Lindbergh and Henry Ford. One essay (429 words long) took the position that Ford's contribution to American life was more important. The other essay (439 words long) took the position that Lindbergh made a more important contribution. Within the bounds of these differences in content, the essays were written so as to be as similar as possible.

Two versions of each essay were made by exchanging synonyms for words in 15 sentences. For example, worthy-deserving, influenced-affected, and wealthiest-richest were three of the synonym pairs. We tested memory for specific words in this experiment (rather than syntactic structure, as in Experiment 1) because we needed to provide the subjects with a comprehensible orienting task. Asking undergraduate subjects to judge word use seemed more likely to be successful than asking them to judge sentence structure.

Two separate tests were created, one for testing content and one for verbatim wording. The content test consisted of 10 true-false questions; each question had opposite answers for the two essays. Thus, prior preference for a particular answer was eliminated as

a possible confound. The questions did not repeat sentences from the essays, but rather stated general arguments or conclusions. The verbatim test consisted of the 15 sentences containing one of the synonyms. The subjects had to identify each sentence as new or old. Because of the differences between the two versions, the old sentences for one version served as the new sentences for the other version. Below each question on both tests was a scale of 1-7, on which the subjects were to indicate their confidence in their answers.

Procedure. The subjects read instructions that informed them that they were to read an essay by a student from a local high school. In the content condition, they were told that they were going to be asked to rate the essay for the quality of its argumentation: "Specifically, you will be asked to rate the consistency of the argument, whether coherent points are made, whether the argument follows a logical progression, and so forth." It was specifically mentioned that differences in writing style and word use were not of interest. In the vocabulary condition, the subjects were told that they were to rate the essay for the level of difficulty and variety of the words used. The instructions specified that differences in argumentation and logic were not of interest.

The subjects read the essays at their own speed. When finished, they evaluated the essay's content or vocabulary with a five-item questionnaire. These data were not analyzed. The subjects then took both the surface and the content memory tests described previously. The order of testing was counterbalanced across subjects and orienting tasks.

Results and Discussion

Four of the subjects, 2 in each group, scored 33% correct in the vocabulary test, which was well below chance. It seemed likely that these subjects were not attending very closely to the materials, so their data were discarded. The mean percent correct of the remaining 28 subjects is shown in Table 2. There was no apparent effect of order of testing, so the data from the counterbalancing groups were collapsed for analysis.

As shown in Table 2, vocabulary instructions improved the accuracy of verbatim memory by about 11%, a significant difference [t(26) = 2.48, p < .02]. In contrast, the content instructions did not improve the accuracy of content memory; the two means were virtually identical [t(26) = .33]. At about 80% accuracy, these levels of performance were both above chance and below ceiling.

The results of this experiment, then, provide partial support for the predictions of the pragmatic view. When the subjects were given an orienting task that directed them to pay attention to wording, they were reliably more accurate in a verbatim memory task. However, instructions to attend to the argument of an essay did not result in improved content memory. It may be that the subjects in both conditions attended to the content at a fairly high level. Although the subjects in the vocabulary condition

Table 2
Percent-Correct Memory for Surface and
Semantic Information in Experiment 2

	Instruction Condition	
	Vocabulary	Logic
Memory for meaning	80	79
Surface memory	66	55

were not required to attend to the arguments of the essay, it may have been difficult not to do so while reading the essay and evaluating its use of vocabulary. It seems possible that a more specific (and more difficult) test of particular kinds of content would show such a difference. This possibility is discussed at greater length in the General Discussion.

EXPERIMENT 3

In the previous experiment as well as in Graesser and Mandler's (1975) study, instructions to get subjects to focus on surface information were used. In some real-life tasks (like editing), people do have an explicit goal to focus on such surface information. But in the normal course of events, listeners focus on the content of a speaker's utterances. Keenan et al.'s (1977) demonstration of good verbatim memory for highly interactive utterances suggests that listeners may themselves vary as to whether they focus on surface information. That is, Keenan et al.'s subjects heard all the sentences as part of a seminar, without any explicit instructions to remember or attend to any part of the talk. Thus, their better memory for some (but not all) sentences suggests that they were actively varying their encoding strategy, depending on the nature of the utterance.

One drawback of Keenan et al.'s (1977) study and those like it is that different items were used in different conditions. For example, the item "Italians, you know what Italians are like, they had a strike, they had a heat wave" was always in the high interactional content condition, whereas "You put a little morpheme that says you're going to choose the object as subject" was always in the low interactional content condition. Because of the nature of the study, it was difficult to ensure that the sentences in different conditions were equivalent in syntax, style, vocabulary, and so forth, and were presented in the same manner. That is, it was difficult to ensure that different sentences were equally memorable, aside from their interactional content. (Actually, Keenan et al. ruled out more of these factors than some subsequent studies, because they showed that high- and low-interactive sentences were remembered equally well in a list-learning task. However, this control study did not rule out the effect of differences in the way the original sentences were presented in the seminar.) Furthermore, because the seminar was (largely) spontaneous speech, it was not possible before the fact to predict which of the sentences would be remembered well and which would be remembered poorly.

Experiment 3 was an attempt to show that subjects can vary how much they encode surface information, without these potential problems. We did this by using the same sentences in two different contexts—one in which they were expected to be highly interactive and one in which they should have low interactiveness. Thus, unlike in Keenan et al.'s (1977) study, items were held constant across conditions. Furthermore, by presenting the text in written form, the physical presentation of the stim-

uli was also constant across conditions. We predicted that by varying the type of text in which the sentences occurred, we could control whether the subjects attended to the surface information in these sentences. Because both contexts were perfectly normal discourses, the integration view would not predict any difference between the conditions—the sentences should be integrated into their contexts in both cases. The processing shift hypothesis also does not seem to predict any difference here (though see the General Discussion). In contrast, according to the pragmatic view, the content of a discourse may cause subjects to focus on surface information, leading to improved memory. In particular, when the sentences were used as criticisms, insults, or other personal remarks, their exact phrasing was crucial to understanding the speaker's intentions. Thus, readers should attend to the wording in this case. If the same sentence was used for a less charged purpose (see examples below), then subjects would attend less to the wording, and verbatim memory should suffer.

Method

Subjects. The subjects were 28 Brown University undergraduates who were paid for participating.

Materials. The two basic materials were each in the form of a letter from a young woman named Samantha. One letter (711 words long) was to her cousin Paul. It was a chatty, very friendly letter describing her baby and rather mundane family outings. The other letter was a highly sarcastic, bitter letter (679 words long) to Samantha's boyfriend, Arthur, reviewing his bad behavior and how much she disliked him. (The subjects only read one letter, so the two were not consistent in content.) We call the first letter the "bland" condition and the second the "sarcastic" condition. There were 10 sentences common to the two letters. These sentences were written so that they were fairly neutral or positive in the bland letter, but would be interpreted as highly sarcastic or insulting in the sarcastic letter. Thus, although the sentences were constant, they would qualify as highly interactive sentences in the sarcastic condition only. For example, the question "Is the flower shop too far for you?" was a sincere question in the bland condition, followed by a suggestion of a different florist. In the sarcastic condition, this same sentence served as a caustic reminder that Arthur had not sent flowers, even though he works for a florist. Another sentence was "It never occurred to me that I would be a mother so young." In the bland condition, this was a literal statement about Samantha's new baby, whereas in the sarcastic condition, it was a commentary on Arthur's juvenile behavior.

For each of the 10 crucial sentences, two different forms that did not differ in meaning were created (e.g., "I never thought I would be a mother at such a young age" for the example given above). These two forms were used to create two versions of each letter.

We constructed a 14-item verbatim recognition test for each letter. It included the 10 crucial items, half in their correct form and half in the alternate form. Two different test forms were used for each letter, so that a given item was old half of the time and new half of the time. There were also four semantically incorrect sentences, which contained words or concepts similar to one of the letters, but which did not correspond to the meaning of any sentence in either letter. These were the same in both forms.

Procedure. The subjects were asked to read one of the letters at their own pace. They were told that they would be asked questions about the letter afterward, but no mention was made of a memory test. After reading the letter, the subjects read a page of in-

structions that explained the verbatim recognition task. The instructions contained a three-sentence passage with sample test items, so that they would understand the task.

The subjects were randomly assigned to letter condition and test form with the restriction of equal numbers of subjects in each condition.

Results and Discussion

Table 3 presents the responses to the three types of test items. As can be seen, the subjects in both conditions were near ceiling for the semantically incorrect items, calling them "old" only about 5% of the time. Thus, the subjects read both letters carefully. However, those who read the sarcastic letter had a higher percentage of hits and a lower percentage of false alarms (to paraphrases) on the surface memory items. The number of hits minus false alarms was calculated for each subject as a measure of verbatim memory. This difference was 43% for the sarcastic letter and 17% for the bland letter, a significant difference [t(26) = 2.76, p < .02].

It should be emphasized that the items on the surface memory test were identical for the two conditions. The same sentences occurred in both letters, but their surface form was much more memorable when they had insulting or sarcastic intent. Thus, the results cannot be attributed to structural or stylistic differences between sentences. Instead, this result suggests that subjects engage in different encoding processes for such sentences, even when they are not given instructions to do so. That is, highly interactive sentences attract more attention to their surface details.

A reviewer pointed out a possible interpretation of the pragmatic view that conflicts with these results. Because the original sentences and surface foils had very similar pragmatic implications, it might seem that subjects should not be able to distinguish them. However, the pragmatic model says that such highly interactive sentences draw attention to their exact wording—not that subjects only remember pragmatic import. The hypothesis that readers remember the speaker's pragmatic intention without remembering the exact wording is certainly a possible view, but the present results suggest that it is incorrect. Instead, they suggest that exact wording is well remembered for these sentences.

As mentioned earlier, neither the integration nor the processing shift view predicts the obtained differences. The sarcastic and insulting sentences were easily integrated into the overall text of the sarcastic letter, as the entire letter was a diatribe against the addressee. Thus,

Table 3
Percent "Old" Judgments to Test Items in Experiment 3

	Story Condition	
	Bland	Sarcastic
Semantic foils	4	5
Hits (old sentences)	71	86
False alarms (paraphrases)	54	43
Hits minus false alarms	17	43

there is no reason that surface memory should be better in that case, according to the integration view. In order for the processing shift view to explain these differences, it would have to be the case that the sarcastic letter had fewer shifts to new structures than the bland letter. This was clearly not the case on a word or sentence level (these were constant in the test items). We do not know for sure that the higher level structures of the letters were equal, but they were certainly not designed to vary in any systematic way. Furthermore, there are questions about whether such higher level structures would influence verbatim memory, as we discuss in the next section.

GENERAL DISCUSSION

The main goal of the experiments described in this article was to investigate the pragmatic view of memory for surface form. Experiment 1 did this indirectly by disconfirming the integration hypothesis, using connected text with subjects who were not expecting a memory test. The results showed that integratable text yielded, if anything, better memory for surface form than scrambled sentences did. Experiment 2 showed that instructions that oriented subjects toward processing surface form improved their memory for that form. And Experiment 3 demonstrated that the nature of the text could itself cause subjects to pay attention to its surface form.

These studies support some crucial components of the pragmatic view. First, the pragmatic view claims that memory for surface form is not qualitatively different from memory for content—they are similarly influenced by the same variables. Thus, the pragmatic view would argue that the effect of integrating text should not have opposite effects on verbatim and content memory (if other variables are held constant). This is exactly what was found in Experiment 1 and under somewhat different conditions by Gernsbacher (1985).

A second claim of the pragmatic view is that surface memory can be improved if subjects actively encode surface information. Because of the function of communication, content is the normal interest of the speaker and understander, and that is the level at which they direct their attention. However, when task demands require attention to surface form, verbatim memory can be improved, as shown in both the present Experiment 2 and Graesser and Mandler (1975). In neither study was it found that instructions improved content memory, however. We have two related explanations for this. One is that the subjects probably never ignored content in these experiments; in fact, the surface task may have required understanding of the content in order to judge the appropriateness of the vocabulary being rated. Also, as is well known, it is very difficult to ignore meaning in meaningful material. The second reason is that it may be misleading to think of "memory for content" as one, unanalyzed whole. In a real story or conversation, the content is a very rich and complex structure, with many aspects. It would be rather unlikely for instructions to

increase or decrease encoding of all this information simultaneously. Instead, it may be that specific themes, ideas, characters, or events are focused on more than others and, therefore, are better remembered. And in fact, such effects within content memory for stories have been found (e.g., R. C. Anderson & Pichert, 1978; Owens, Bower, & Black, 1979). On this view, it would not be possible to give instructions that would cause subjects to have an overall loss of content memory, unless the task discouraged reading for meaning at all.

The final component of the pragmatic model that was demonstrated here was that text can demand attentional resources that influence memory for verbatim material. As Keenan et al. (1977) showed, some kinds of sentences are well remembered for a few days after being heard, whereas the surface characteristics of other kinds are forgotten in the same period. Experiment 3 showed that this result is not due to differences in presentation, wording, syntactic structure, or any sentence-level structures. Rather, it is the interpersonal characteristic of the highly interactive sentences that leads to their better memory. When these sentences appeared as normal, mundane speech acts, their surface representations were not remembered very well (a 17% difference between hits and false alarms). When the same sentences appeared in contexts that made them seem sarcastic, insulting, or emotionally laden, they were remembered very well (a 46% difference). This result demonstrates that readers may allocate their attention to different levels of analysis in a discourse, depending on its pragmatic import. Because the exact wording of an insult or a compliment is crucial to understanding its impact, subjects attend much more closely to the wording of such sentences. Because the precise form of a neutral question or statement is not so crucial, subjects do not attend to the surface form, and so have poor memory for it. MacWhinney et al. (1982) showed that the difference between these two forms is not due to highly interactive sentences increasing arousal and, therefore, memory.

The three experiments reported here have focused on the importance of encoding processes in determining the level of verbatim memory (and we assume that retrieval processes also have important effects; see Morris et al., 1977; Roediger & Blaxton, 1987). A strength of the pragmatic view is that it ties this memory theory to assumptions about how readers normally encode text. We have focused on the differences between sentences with high and low interactiveness, because they have a particular influence on verbatim memory. However, this view could be broadened to study other aspects of comprehension that receive different kinds of processing according to some theory, in order to discover whether their memory representation differs accordingly. In fact, the work of Gibbs (1981) and Murphy (1992) can be interpreted in just this light. Gibbs showed that the surface forms of conventional requests were not remembered as well as those of unconventional requests. One interpretation of this is that unconventional request forms draw attention to their exact wording, whereas conventional forms can be interpreted without much analysis of the way they are phrased. Similarly, Murphy (1992) argued that the social information included in personal terms of address should cause them to be remembered better than object names, and his results did find better verbatim memory for such terms.

Finally, it should be noted that other variables familiar to memory researchers may well be partly responsible for the generally poor memory of verbatim material. For example, there may be greater similarity of verbatim information in a text, which would cause increased interference of verbatim information. Common words and their synonyms may be repeated, syntactic structures recur, and intonation patterns repeat numerous times in a conversation. For example, in this article, we have used the phrases "verbatim memory," "surface memory," and "memory for form" somewhat interchangeably. As a result, interference would make it difficult for a reader to recognize which of these phrases occurred in a given sentence. Thus, there may be a variety of other factors that also contribute to poor verbatim memory (or poor surface memory [or poor memory for form]).

In this vein, Gernsbacher's (1985) processing shift hypothesis may well account for part of the poor quality of verbatim memory. Experiments 2 and 3 do not in any way disconfirm this view, but we have argued that the processing shift hypothesis simply does not predict these effects (nor do Keenan et al.'s, 1977, findings). But, by the same token. Gernsbacher's finding that surface memory is worse right after a new structure boundary is crossed is an important one (see also Clark & Sengul, 1979; Jarvella, 1979) that the pragmatic view does not account for. Thus, there may well be some truth to the processing shift hypothesis, independent of the explanation we have given. It may be that surface memory is a function both of the structure of the discourse, as proposed by the processing shift hypothesis, and of the level of encoding, as proposed by the pragmatic view. We believe that processing shifts at lower levels of analysis (phrases, clauses, sentences) are more likely to influence verbatim memory than those at higher levels (discourse topics, story episodes, etc.), because higher level discourse processing seems unlikely to directly modify or even activate the representation of surface form. That is, beginning a new clause may well divert processing from the words and syntactic structure of the previous clause, but it seems unlikely that beginning a new story episode will have an effect on surface information of the entire previous episode. At least, any such effects should be just as pronounced in memory for content of the prior episode.

The main goal of the present experiments, however, was to emphasize the importance of nonstructural factors on verbatim memory, and especially the way that highly interactive sentences can attract attention to their precise wording. The experiments presented here demonstrate that

such factors do influence memory for surface information, and so any account of verbatim memory will have to take these pragmatic variables into account.

REFERENCES

- ANDERSON, J. R. (1974). Verbatim and propositional representation of sentences in immediate and long-term memory. *Journal of Verbal Learning & Verbal Behavior*, 13, 149-162.
- ANDERSON, J. R., & BOWER, G. H. (1973). Human associative memory. Washington, DC: Winston & Sons.
- ANDERSON, J. R., & PAULSON, R. (1977). Representation and retention of verbatim information. *Journal of Verbal Learning & Verbal Behavior*, 16, 439-451.
- ANDERSON, R. C., & PICHERT, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning & Verbal Behavior*, 17, 1-12.
- BARTLETT, F. C. (1932). Remembering: A study in experimental and social psychology. Cambridge: Cambridge University Press.
- BIRCH, S. (1993). Focus effects in sentence comprehension. Unpublished doctoral dissertation, University of Illinois.
- BOCK, J. K., & BREWER, W. F. (1974). Reconstructive recall in sentences with alternative surface structures. *Journal of Experimental Psychology*, 103, 837-843.
- Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. Cognitive Psychology, 11, 177-220.
- BRANSFORD, J. D., BARCLAY, R. J., & FRANKS, J. J. (1972). Sentence memory: A constructive versus interpretive approach. Cognitive Psychology, 3, 193-209.
- Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology*, 2, 331-350.
- Bransford, J. D., & Johnson, M. K. (1973). Considerations of some problems of comprehension. In W. G. Chase (Ed.), *Visual information processing* (pp. 383-438). New York: Academic Press.
- CLARK, H. H., & SENGUL, C. J. (1979). In search of referents for nouns and pronouns. Memory & Cognition, 7, 35-41.
- CRAIK, F. I. M., & LOCKHART, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Ver*bal Behavior, 12, 599-607.
- CRAIK, F. I. M., & TULVING, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psy*chology: General, 104, 268-294.
- Gernsbacher, M. A. (1985). Surface information loss in comprehension. Cognitive Psychology, 17, 324-363.
- Gibbs, R. W., Jr. (1981). Memory for requests in conversation. Journal of Verbal Learning & Verbal Behavior, 20, 630-640.
- GIBBS, R. W., JR., & NAGAOKA, A. (1985). Getting the hang of American slang: Studies on understanding and remembering slang metaphors. Language & Speech, 28, 182-194.
- Graesser, A., II, & Mandler, G. (1975). Recognition memory for the meaning and surface structure of sentences. *Journal of Experimental Psychology: Human Learning & Memory*, 1, 238-248.
- JARVELLA, R. J. (1971). Syntactic processing of connected speech. Journal of Verbal Learning & Verbal Behavior, 10, 409-416.
- JARVELLA, R. J. (1979). Immediate memory and discourse processing. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 13, pp. 379-421). New York: Academic Press.
- JOHNSON-LAIRD, P. N., & STEVENSON, R. (1970). Memory for syntax. *Nature*, 227, 412.

- KEENAN, J. M., MACWHINNEY, B., & MAYHEW, D. (1977). Pragmatics in memory: A study of natural conversation. *Journal of Verbal Learning & Verbal Behavior*, 16, 549-560.
- KINTSCH, W., & BATES, E. (1977). Recognition memory for statements from a classroom lecture. *Journal of Experimental Psychology: Hu*man Learning & Memory, 3, 150-159.
- KOLERS, P. A. (1975). Memorial consequences of automized encoding. *Journal of Experimental Psychology: Human Learning & Memory*, 1, 689-701.
- Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology*, 14, 78-106.
- MACWHINNEY, B., KEENAN, J. M., & REINKE, P. (1982). The role of arousal in memory for conversation. *Memory & Cognition*, 10, 308-317.
- MALT, B. C. (1985). The role of discourse structure in understanding anaphora. *Journal of Memory & Language*, 24, 271-289.
- Mani, K., & Johnson-Laird, P. N. (1982). The mental representation of spatial descriptions. *Memory & Cognition*, 10, 181-187.
- MORRIS, C. D., BRANSFORD, J. D., & FRANKS, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior*, 16, 519-533.
- MURPHY, G. L. (1985a). Processes of understanding anaphora. *Journal of Memory & Language*, 24, 290-303.
- MURPHY, G. L. (1985b). Psychological explanations of deep and surface anaphora. *Journal of Pragmatics*, 9, 785-813.
- MURPHY, G. L. (1992). Comprehension and memory for personal reference: The use of social information in language processing. *Discourse Processes*, 15, 337-356.
- Noice, H. (1992). Elaborative memory strategies of professional actors. *Applied Cognitive Psychology*, **6**, 417-427.
- OWENS, J., BOWER, G. H., & BLACK, J. B. (1979). The "soap opera" effect in story recall. Memory & Cognition, 7, 185-191.
- PETERSON, R. G., & McINTYRE, C. W. (1973). The influence of semantic "relatedness" on linguistic integration and retention. *American Journal of Psychology*, **86**, 697-706.
- POMPI, K. F., & LACHMAN, R. (1967). Surrogate processes in the short-term retention of connected discourse. *Journal of Experimental Psychology*, 75, 143-150.
- ROEDIGER, H. L., III, & BLAXTON, T. A. (1987). Effects of varying modality, surface features, and retention interval on priming in word-fragment completion. *Memory & Cognition*, 15, 379-388.
- SACHS, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, 2, 437-442
- SACHS, J. S. (1974). Memory in reading and listening to discourse. Memory & Cognition, 2, 95-100.
- SAG, I. A., & HANKAMER, J. (1984). Toward a theory of anaphoric processing. Linguistics & Philosophy, 7, 325-345.

NOTE

1. We do not report this experiment in detail here, because it did not produce a reliable advantage of content memory, due to ceiling effects. Nonetheless, it showed a nonsignificant decrease in *both* content and surface memory as a result of nonintegratable text. Thus, it is another disconfirmation of the integration view's prediction that integration will hurt surface memory.

(Manuscript received February 18, 1993; revision accepted for publication May 12, 1993.)