

Pictures in Sentences: Understanding Without Words

Mary C. Potter
Massachusetts Institute of Technology

Judith F. Kroll
Mount Holyoke College

Betsy Yachzel, Elisabeth Carpenter, and Janet Sherman
Massachusetts Institute of Technology

SUMMARY

To understand a sentence, the meanings of the words in the sentence must be retrieved and combined. Are these meanings represented within the language system (the lexical hypothesis) or are they represented in a general conceptual system that is not restricted to language (the conceptual hypothesis)? To evaluate these hypotheses, sentences were presented in which a pictured object replaced a word (rebus sentences). Previous research has shown that isolated pictures and words are processed equally rapidly in conceptual tasks, but that pictures are markedly slower than words in tasks requiring lexical access. The lexical hypothesis would therefore lead one to expect that rebus sentences will be relatively difficult, whereas the conceptual hypothesis would predict that rebus sentences would be rather easy.

Sentences were shown using rapid serial visual presentation (RSVP) at a rate of 10 or 12 words per second. With one set of materials (Experiments 1 and 2), readers took longer to judge the plausibility of rebus sentences than all-word sentences, although the accuracy of judgment and of recall were similar for the two formats. With two new sets of materials (Experiments 3 and 5), rebus and all-word sentences were virtually equivalent except in one circumstance: when a picture replaced the noun in a familiar phrase such as *seedless grapes*. In contrast, when the task required overt naming of the rebus picture in a sentence context, latency to name the picture was markedly longer than to name the corresponding word, and the appropriateness of the sentence context affected picture naming but not word naming (Experiment 4).

The results fail to support theories that place word meanings in a specialized lexical entry. Instead, the results suggest that the lexical representation of a noun or familiar noun phrase provides a pointer to a nonlinguistic conceptual system, and it is in that system that the meaning of a sentence is constructed.

In what part or module of the cognitive system is the semantic information about a noun stored? Is it stored in a lexicon that is part of a linguistic system, or is the meaning simply a part of a general-purpose conceptual system? These two theoretical possibilities place the division between linguistic and nonlinguistic thought at fundamentally different points. The contrast can be highlighted by considering two ways in which the meanings of words might be put together in sentence processing. According to one approach, a word's representation in the lexicon provides only a pointer to the relevant concept, and the composition of word meanings occurs in a general-purpose conceptual system. In this view, a person reading or listening to a sentence activates conceptual information about each content word and builds a representation of the sentence's meaning in the conceptual system. According to the other approach, core semantic information is included in the lexical representation itself. The lexical representation is one component of the linguistic system, which is a modular processor separate from the general-purpose conceptual system. Sentence comprehension is viewed as including two distinct although possibly overlapping stages: semantic composition, which takes place in the linguistic sys-

tem and yields the sentence's literal meaning; and pragmatic interpretation of this meaning in the general-purpose conceptual system.

Although many cognitive psychologists find the conceptual approach congenial, linguists and psycholinguists have focused on the linguistic system, and for the most part they explicitly or implicitly accept some version of the lexical approach. One reason is that language-specific lexical representations rather than the underlying concepts are involved in some syntactic constraints. For example, *Mädchen* is the German word for a young woman, but it is neuter and takes the neuter article *das*, the anaphoric pronoun *es*, and so forth. In English, the singular concept *scissors* has a plural name. Gender and number agreement in such cases is determined by the lexical form rather than the concept. Another example is verb subcategorization; for instance, *eat* can be used both transitively and intransitively (*He ate*), but *devour* can only be used transitively (**He devoured*). These differences in syntax seem to have little or nothing to do with differences in the concepts of the contrasting verbs. Such examples show that the lexical representation is not simply replaced by a nonlinguistic concept during

comprehension, at least not before the syntactic analysis has been completed.

A division between information in the lexicon and in a general purpose conceptual system is also motivated by semantic theory. Some aspects of sentence meaning seem to follow directly from word meanings, whereas others depend on facts about the world. The classic distinction between anomaly and falsehood is a case in point. *My sister is married to a bachelor* is anomalous because of the contradiction between the meanings of *married* and *bachelor*, whereas *My sister is married to Henry VIII* is factually false. Implications of sentences also seem to fall into two categories: those entailments that follow from word meanings (*He was murdered* implies that he is dead) and those inferences that are based on general knowledge (*He was born in 1600* implies that he is dead). These and other observations suggest that some core features of meaning are represented in the lexicon and are used in arriving at a literal interpretation of the sentence, prior to or independent of retrieval of general purpose knowledge (for reviews, see Akmajian, Demers, & Harnish, 1979, Clark & Clark, 1977, and J. D. Fodor, 1977).

This study is an initial attempt to distinguish empirically between the conceptual and lexical approaches to the processing of word meanings. The method we used was to present written sentences in which pictures replaced one or two concrete nouns (rebus sentences). The reason for using pictures as word substitutes is that words and pictures, when presented in isolation, have the following two properties.

1. Written words are named more than 200 ms faster than matched pictures (e.g., Cattell, 1886; Fraisse, 1960; Paivio, 1971, 1978; Potter & Faulconer, 1975). Naming latency is an index of relative time to access a lexical representation (Forster, 1981; Frederiksen & Kroll, 1976) and thus the extra 200 ms required to name a picture indicates a 200-ms delay in retrieving the appropriate lexical entry.

2. The same pictures and words are understood equally fast (if anything, pictures are faster) in a variety of tasks such as categorizing the items or judging their relevance to a preceding sentence (e.g., Banks & Flora, 1977; Potter, 1979; Potter & Faulconer, 1975; Potter, So, Von Eckardt, & Feldman, 1984; Potter, Valian, & Faulconer, 1977; Snodgrass, 1980, 1984).

These findings indicate that pictured objects and the corresponding words share a common conceptual representation that is separate from the lexicon itself, and that written words and pictures access this conceptual representation equally rapidly.

Turning now to sentence comprehension, if the lexical entry

contains semantic information that is used in arriving at an initial interpretation of the sentence (core information that is distinct from that in the all-purpose conceptual representation), then encountering a rebus picture should impose a delay in sentence processing of about 200 ms. On the other hand, if the meaning of a word is represented entirely in the all-purpose conceptual system, a picture would provide equally rapid access to that information, and there should be no particular difficulty or delay in understanding a rebus sentence. (This prediction rests on the strong assumption that the strange look of a rebus sentence, the difficulty of recognizing mixtures of pictures and words, and other similar factors would not produce disruptions or delays in processing. Such effects, if present, would bias the results in favor of the lexical hypothesis.)

To make the task of reading and responding to rebus sentences and to matched all-word sentences sufficiently difficult and time constrained to reveal any disruptive effect of the rebus pictures, rapid serial visual presentation (RSVP) was used (Forster, 1970; see Potter, 1984, for a review). In RSVP, each word of a sentence appears successively at the same location, so that no eye movements are needed and the rate of reading is under experimenter control.

To summarize the logic of the first experiment, if the lexical approach is correct, the lexical representation of a pictured object would have to be retrieved in order to fit the word substitute into the sentence. Because lexical retrieval is substantially slower for pictures than for written words, a marked disruption of sentence processing would be expected under the time-limiting conditions of RSVP reading. If, however, the semantic component of the lexicon consists simply of a pointer to the conceptual system, then rebus sentences should be readily understood, because an appropriate picture would point to the same concept as the noun it replaced.

Experiment 1

To assess sentence processing when a picture replaces a noun, two responses were studied: (a) a speeded decision about plausibility that required comprehension of the sentence (e.g., Levelt & Kempen, 1975), and (b) immediate recall of the sentence. Because previous work (e.g., Aaronson, 1976; Green, 1977) had shown that readers may adopt different strategies for comprehension and for recall, three groups of subjects were compared: a comprehension-only group, a recall-only group, and a comprehension-plus-recall group. The comprehension-only group made plausibility decisions, the recall-only group wrote down the sentence, and the comprehension-recall group did both. The presence or absence of a picture in the sentence, the plausibility of the sentence, and the length of the sentence were varied within subjects.

The rate of presentation was set at 12 words per second because pilot work showed that RSVP reading at that rate was possible but moderately difficult. The rate (equivalent to 720 words per minute) is more than twice that of typical college readers.

Method

Subjects. Forty college-age men and women volunteers were paid for participating in the experiment; 16 were assigned to the comprehen-

This research was supported in part by National Science Foundation Grants BNS77-2-5543, BNS80-2-4453, and BNS83-18156 to the first author, and by Defense Advanced Research Projects Agency Contract MDA903-76-0441 to the first and second authors. We thank Virginia Valian, Janellen Huttenlocher, Susan Lima, Gay Snodgrass, Irvin Rock, and Robert Welker for their comments and Linda Lombardi for assisting in the research.

Correspondence concerning this article should be addressed to Mary C. Potter, Department of Brain and Cognitive Sciences, E10-032, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

sion-recall group, 8 to the comprehension-only group, and 16 to the recall-only group.

Materials. The main experiment consisted of 32 sentences, 8 sentences of each of four lengths, 8, 10, 12, or 14 words. The sentences varied widely in subject matter and grammatical structure; all included at least one concrete, picturable noun, whose serial position varied. There were four versions of each sentence; format (rebus vs. all-word) was crossed with plausibility. Sentences were made implausible by changing one or two words, usually the last word in the sentence. In all cases, sentence plausibility hinged on the last word. For instance, in the implausible version *moon* replaced *lightbulb* in the following sentence: *Judy needed the stool to reach the lightbulb*. In the rebus version of each sentence, *stool* was replaced by a picture of a stool. Other examples are given in the Appendix.

Eight additional sentences were intermixed with the 32 main sentences. All were 12 words in length and half were plausible, half implausible. They were designed to assess the effect of presenting 0, 1, or 2 pictures, so each sentence included at least two picturable nouns. In different versions, both, one, or neither word was replaced by a picture. There were 6 practice sentences.

The pictures used in the rebus sentences were line drawings from a larger set used by Potter and Faulconer (1975). The sentences were typed in lowercase letters and then photographed, one word or picture to a frame, on 16-mm high-contrast double negative film, so that the pictures and words were white on a gray background. The subjects sat 3 m from the screen; a seven-letter word and the largest dimension of the picture each subtended about 4.4°. A warning signal consisting of a row of asterisks appeared for 83 ms, 333 ms before each sentence.

The rationale for the experiment depended on the assumption that, in isolation, the pictures to be used in the rebus sentence would take longer to name than the corresponding words, but would be as easy to perceive and understand as the words. Pretests of the rebus pictures and corresponding words were carried out to test those assumptions. There were two groups of 16 subjects each. In the first group, naming latency for the pictures and words was measured. A warning row of asterisks preceded the stimulus item by 500 ms; the item appeared for 83 ms, preceded and followed by a row of symbols (to mimic the masking effect of RSVP words in the main experiment). In the second group, time to understand a word or picture was assessed using the category-matching task of Potter and Faulconer (1975) and others. A written superordinate category name was presented 667 ms before the target word or picture (which was masked as in the naming condition), and the subject's task was to decide whether the object named or pictured was a member of that category.

In the naming group, subjects were instructed to name the word or picture as rapidly as possible; a voice key measured reaction time (RT) from the onset of the stimulus. In the category-matching group, subjects were instructed to press one response key if the word or picture referred to a member of the specified superordinate category (e.g., *furniture*; *gardening equipment*), and to press the other if it did not. The category and

item matched on half the trials. Match-mismatch was counterbalanced across films, so that each item was seen once in each of the four combinations of match-mismatch and picture-word form.

The results of the pretests, shown in Table 1, replicate those of Potter and Faulconer (1975) in all important respects. In the naming task, the word advantage of 218 ms was significant, $t(1, 15) = 14.17, p < .01$. Although there were 9% errors in picture naming, only 3% were total misunderstandings of the picture; the rest were semantically close responses, such as *car* for *bus*. In the category-matching task, there was no significant difference between pictures and words in RTs or errors. These pretest results show that the critical pictures and words met the requirements for use in Experiment 1: In isolation, the pictures to be used in rebus sentences took substantially longer to name than the corresponding words but were understood just as rapidly and accurately.

Equipment. A 16-mm variable-speed projector was used to present the stimuli. A white transparent square appeared in the lower left corner of the frame with the last word of each sentence, and a photocell activated by the light spot started a pair of clock counters. The subject pressed one of two response buttons to indicate whether the sentence was plausible; RT was measured to the nearest millisecond.

Design and procedure. Each subject saw a set of sentences in which all-word and rebus formats and plausible and implausible sentences appeared equally often, counterbalanced over four sentence lengths. There were two blocks, each consisting of 16 sentences (plus fillers), which comprised a complete replication of the two formats, two plausibilities, and four lengths. The order of sentences was randomized within block and that same order was used in all conditions. The four versions of each sentence were counterbalanced over four films, each seen by a quarter of the subjects in each of the groups.

The comprehension-recall group first decided whether the sentence was plausible (by pressing one of two keys) and then wrote it down. The comprehension-only group and the recall-only group performed one or the other task, respectively. All groups were told that some sentences would be plausible and others implausible and that pictures would replace words in some sentences.

Results

In initial analyses, the comprehension-recall group's results were compared with the recall-only group (for sentence recall) and with the comprehension-only group (for plausibility judgments). In most comparisons there were no differences between groups, so only the combined results are reported except when group differences were found. In no case did the rebus variable interact with group. A breakdown of the results by groups is shown in Table 2 (for the plausibility judgment) and Figure 1 (for recall). To summarize the overall results briefly, subjects took 1,345 ms to make a decision about the plausibility of sentences and made .12 errors (weighted mean of the two groups).

Table 1
Pretest for Experiment 1: Mean Reaction Times (in Milliseconds), Standard Deviations, and Error Rates in Naming or Categorizing Pictures and Written Words

Group	Pictures			Words			Word advantage	
	RT	SD	E	RT	SD	E	RT	E
Naming	752	61	.09	534	67	.01	218	.08
Category-matching								
Yes	648	118	.05	646	108	.12	2	-.06
No	652	114	.09	668	168	.07	-16	.02
M	650		.07	657		.09	-7	-.02

Table 2
 Experiments 1 and 2: Mean Reaction Times (in Milliseconds), Standard Deviations, and Error Rates
 in Judging the Plausibility of Rebus and All-Word Sentences

Group	Plausible						Implausible					
	Rebus			Words			Rebus			Words		
	RT	SD	E	RT	SD	E	RT	SD	E	RT	SD	E
Experiment 1												
Comprehension-recall	1,310	847	.10	1,261	508	.14	1,520	676	.17	1,392	623	.14
Comprehension only	1,262	463	.06	1,133	342	.06	1,457	539	.11	1,326	494	.08
Experiment 2												
Comprehension-recall	1,364	626	.09	1,233	476	.07	1,529	591	.18	1,424	651	.10

In immediately recalling the sentences, 14% of the words were omitted or recalled incorrectly. (In most cases, the gist of the sentence was recalled correctly; more will be said about the nature of the recall errors in the final discussion.) Therefore, as intended, reading was moderately difficult but not dramatically impaired. Rebus sentences could be understood and recalled almost—but not quite—as easily as all-word sentences. Recall errors were slightly higher (by 1.4%) for the rebus sentences, but the difference was not significant. It took subjects 103 ms longer to judge the plausibility of rebus sentences than all-word sentences (a difference not obtained in Experiments 3 and 5 with new sets of sentences), but there was no difference in the accuracy of the decision, showing that a picture does provide the information necessary to understand a sentence in which it appears. Details of these and other analyses follow.

Analyses of variance (ANOVAS) were carried out for subjects (F_1) on the means of their responses to the two sentences of the same length, format, and plausibility, and for sentences (F_2) on the means of the subjects who saw a given sentence in the same format and plausibility condition. For the RT analyses, only correct responses were included; empty cells were replaced using Winer's procedure (Winer, 1962). Correct responses greater than two standard deviations (SD s) from a subject's mean RT (4% of all responses, in Experiment 1) were truncated to that score.

Plausibility. The mean RTs and error rates for the comprehension-recall and comprehension-only groups are shown in Table 2: as noted, the groups did not differ significantly; therefore, only analyses for the combined groups are reported. Analyses of the plausibility-decision errors, using a t test of differences and (for length) the Friedman two-way ANOVA, showed no significant effects of format, plausibility, or sentence length. In the analysis of decision time, all-word sentences were responded to 103 ms faster than rebus sentences, $F'_{min}(1, 50) = 4.29, p < .05$. The decision that a sentence was plausible was made 178 ms faster than the decision that it was implausible, $F'_{min}(1, 49) = 4.59, p < .05$; there was no interaction with format. The length of a sentence did not have a significant effect on decision time; the mean RTs (with error rates in parentheses) for sentences of 8, 10, 12, and 14 words respectively, were 1,299 ms (.09), 1,382 ms (.13), 1,372 ms (.12) and 1,328 ms (.14), and there was again no interaction with rebus versus all-word

format. The fact that decision time did not increase systematically with increasing length of sentence (at least for lengths 10–14) suggests that subjects were able to process sentence meaning on line, that is, while the words and picture were being presented.

Omissions in recall. Recall results are shown in Figure 1. ANOVAS were carried out on the arcsine-transformed proportion of omitted words per sentence. There was no significant difference between comprehension-recall and recall-only groups in the subjects analysis, $F_1(1, 30) = 0.75$, although there was a significant difference (favoring the comprehension-recall group) in the items analysis, $F_2(1, 28) = 7.16, p < .025$. Sentence format had no significant effect on recall: 13% of the words were omitted in all-word sentences and 14% were omitted in rebus sentences. The rebus picture itself was incorrectly identified or omitted in only 7% of the sentences, and the corresponding word was missed in 6% of the all-word sentences. Nor did plausibility significantly affect recall of the sentence; there were 13% omissions in plausible sentences and 14% in implausible sentences.

The length of the sentence, however, had a significant effect

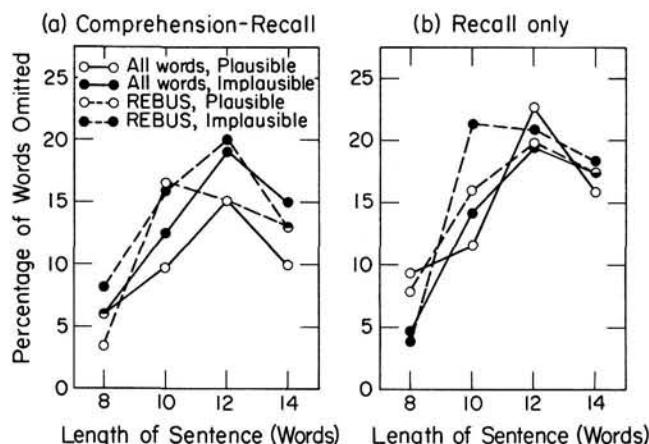


Figure 1. Experiment 1: The percentage of words omitted in recall of sentences in (a) the comprehension-recall group and (b) the recall-only group.

on recall, $F'_{\min}(3, 34) = 5.63, p < .01$ (see Figure 1). A Newman-Keuls test showed that the percentage of words omitted was lowest for 8-word sentences (6%), highest for 12-word sentences (19%), and intermediate for 10-word and 14-word sentences (15% each). The low error rate for sentences as short as 8 words is not particularly surprising. The high proportion of errors on the 12-word sentences seems to have been a sampling effect, because a different group of 12-word sentences (see *Two pictures in a sentence*, later) had 15% errors. Thus, the error rate per word did not increase systematically as the number of words in a sentence increased from 10 to 14. There was a marginally significant three-way interaction of sentence length with group and plausibility (see Figure 1), $F_1(3, 90) = 2.29, p < .10$; $F_2(3, 28) = 4.08, p < .05$. Plausible sentences were recalled more accurately than implausible sentences by the comprehension-recall group, at each sentence length, whereas the recall-only group showed no consistent effect of plausibility on recall. No other interactions were significant.

Two pictures in a sentence. In eight extra 12-word sentences (mixed with the main set of sentences), the effect of presenting two pictures, one or the other picture, or no pictures in a given sentence was assessed (each of the eight sentences appeared in all four forms, counterbalanced across subjects). The question was whether any problems encountered with a single picture would be exaggerated when there were two pictures. For example, if a rebus picture were processed separately from the rest of the sentence and then fitted in (as might happen if a picture's name had to be retrieved), two rebus pictures might be even more difficult to manage. Subjects took 1,179 ms (with .10 errors) to decide the plausibility of all-word sentences, 1,280 ms (with .11 errors) when there was one or the other picture, and 1,243 ms (with .10 errors) when there were two pictures. Thus, as in the main results, sentences with any pictures were responded to somewhat more slowly than all-word sentences, $t(23) = 1.48, p < .10$, one-tailed, although just as accurately. In recall of the sentences, 16% of the words in all-word sentences were omitted, compared with 14% in one-picture sentences and 14% in two-picture sentences: Clearly there was no impairment in recall due to pictures. In sum, there was no hint of a further increase in the difficulty of comprehending and recalling a sentence when the number of pictures increased from one to two.

Discussion

The presentation of sentences serially at a rate of 12 words per second succeeded in taxing the ability of subjects to understand and report the sentences, even when no picture was included. That is most clearly shown by the recall results (Figure 1), in which the proportion of errors would have been near zero if the sentences had been presented at a slower rate. Plausibility-judgment errors were also sufficiently high (12% overall) to assure that ceiling effects would not obscure possible difficulties introduced by a rebus picture. At the same time, most subjects on most trials did evaluate plausibility correctly and did recall at least the gist of the sentence, suggesting that the outcome of sentence processing in the present tasks was not drastically different from normal.

Under these fairly difficult reading conditions, a marked disruption in processing rebus sentences should have resulted if a

picture had not provided the information needed to fit it into the sentence. Instead, rebus sentences were understood and recalled as accurately as all-word sentences, although the plausibility judgment took 103 ms longer. (As already mentioned, this RT difference was not replicated in later experiments that used different materials.) Given the 218-ms disparity between picture- and word-naming time in the pretest, the lexical representation of the rebus picture should have arrived belatedly and probably out of order. When reading so rapidly, subjects would have little opportunity to recover from such a delay; one might have expected substantial disruption of recall (as in Mitchell, 1979) and a marked increase in mistaken plausibility judgments, not just an increase in RT that was half the magnitude of the disparity in naming latency. Therefore, it is reasonable to conclude that information stored with or accessed exclusively from a lexical entry is not essential for sentence processing. A conceptual representation (readily available from the picture) could be integrated rather smoothly into the sentence.

In Experiment 1, it was not only the lexical status of a rebus picture that made it different from the word it replaced; surface characteristics such as global shape and size also made the picture distinctive. Experiment 2 investigated the possibility that the longer RTs and slightly lower recall accuracy for rebus sentences observed in Experiment 1 could be due simply to the startle effect of a shift in appearance. On the other hand, the distinctiveness of the picture format might have helped the viewer to pick out that important "word" in the sentence, leading to an underestimation of the difficulty of rebus pictures (Theios & Freedman, 1984, have shown that large-sized pictures have an advantage over smaller sized words). Experiment 2 permitted us to test this hypothesis as well. In Experiment 2's all-word sentences, the word corresponding to the rebus picture was made visually surprising by changing it to uppercase letters and doubling its size. Thus, both the rebus picture and the critical word in all-word sentences were visually distinctive.

Experiment 2

Method

Subjects. Sixteen new subjects from the same pool as those in Experiment 1 were paid for their participation.

Materials. The sentences were identical to those of Experiment 1, except that only two of the four sets of materials were used. Format (rebus picture vs. one large word) was counterbalanced over individual sentences but plausibility was not; a fixed half of the sentences were plausible. In the all-word sentences, the words corresponding to rebus pictures were printed in block uppercase letters about twice the width and height of those used for the other words of the sentence. In viewing the sentence, that word seemed to expand or pop outward, just as a rebus picture gave the impression of popping outward.

Procedure. The procedure was like that of the comprehension-recall group in Experiment 1 except that 8 subjects saw one of the two sets of materials, 8 subjects saw the other, and all subjects were told that sentences might include a large word or picture.

Results

The main question of Experiment 2 was whether large words would eliminate (or exaggerate) the reaction time difference between all-word and rebus sentences: They did neither. The over-

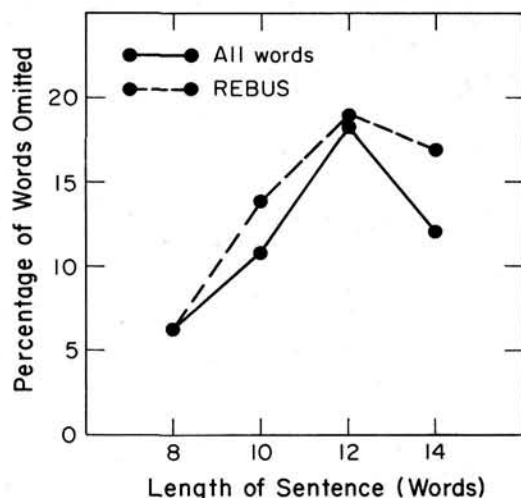


Figure 2. Experiment 2: The percentage of words omitted in recall.

all word advantage observed in Experiment 1 was not eliminated. The same pattern of results—longer RTs and slightly lower recall for rebus sentences—was observed in Experiment 2, although the differences were not significant. The details of these and other analyses follow.

Unlike Experiment 1, the number of errors in making the plausibility decision (Table 2) was higher for rebus sentences than all-word sentences, according to a *t* test of differences, $t(15) = 2.16$, $p < .05$, two-tailed. There were also more errors on implausible than plausible sentences, $t(15) = 2.39$, $p < .05$; this effect did not interact with sentence format. All-word sentences were 118 ms faster than rebus sentences, overall, but the difference was reliable in neither the subjects nor the sentences analysis, $F_1(1, 15) = 3.07$, $p = .10$; $F_2(1, 24) = 2.84$, $p > .10$. The only variable that significantly affected response time was plausibility, $F'_{\min}(1, 38) = 4.16$, $p < .05$, with plausible sentences 179 ms faster than implausible ones.

Figure 2 shows the percentage of words omitted in recall. Analyses were carried out on the arcsine-transformed proportion of errors per word. The main effect of format (rebus vs. words) was not significant. There was a slight advantage of plausible sentences in the subjects analysis but not the sentences analysis, $F_1(1, 15) = 6.15$, $p < .05$, $F_2(1, 24) = .78$; an interaction with format was significant only in the subjects analysis ($p < .05$). Length of sentence was significant, $F'_{\min}(3, 32) = 4.92$, $p < .01$, with a pattern similar to that of Experiment 1: That is, 8-word sentences were more accurately recalled than longer sentences, and 12-word sentences were less accurately recalled than either 10- or 14-word sentences (Figure 2).

Analyses were carried out by combining and comparing Experiment 1's comprehension-recall group and the subjects in Experiment 2. In none of the analyses was there a significant difference between the experiments; all *F*s were less than 1. In particular, for the all-word sentences, there was no difference between Experiments 1 and 2. The results were consistent with those already reported and will not be presented in detail. It is worth noting that with a combined $N = 32$, the main difference between rebus sentences and all-word sentences, namely, the

104-ms advantage of the latter in the plausibility decision, was significant in the subjects analysis ($p < .025$) but was only marginally significant in the sentences analysis ($p < .10$). Similarly, the small increase in recall errors with rebus sentences (13.6% vs. 11.9%) was significant in the subjects analysis ($p < .025$) but not the sentences analysis ($p > .25$). In contrast, the effect of sentence length on the accuracy of recall was significant by $F'_{\min}(p < .01)$, suggesting that the marginality of the rebus effect was not simply due to insensitivity.

Discussion

If the small rebus disadvantage observed in Experiment 1 had been a result of the startling appearance of a picture in the sentence, an enlarged word might have been expected to produce the same effect. However, the overall word advantage in time to judge plausibility (and the slight advantage in recall) persisted in Experiment 2. (The word advantage did not increase, as would have been expected if the distinctiveness of an important content word actually helps processing.)

Still, the main conclusion to be drawn from Experiment 2, like Experiment 1, is that pictures can be understood remarkably readily even in sentences presented sequentially at 12 words per second. This seemingly direct incorporation of the picture concept into the sentence as it is being read supports the hypothesis that word meaning is not represented as part of a purely lexical representation, but rather is represented in a nonlexical, conceptual system.

Experiment 3

Experiment 3 had two goals. One was to replicate Experiments 1 and 2 with new materials. The other was to place the critical picture or word at the end of the sentence. In the previous experiments the rebus picture almost always appeared before the end of the sentence. We assumed that subjects would be obliged to deal with it immediately in order not to miss the rest of the sentence, and that only the pictured concept, not on a more slowly retrieved lexical representation, would be available for immediate processing. Still, there remained a possibility that subjects had time, before the end of the sentence, to retrieve a lexical representation for the picture and fit in the relevant semantic information.

In Experiment 3, therefore, the critical word or picture always appeared at the end of the sentence, and it determined whether or not the sentence was plausible. To avoid an implicit demand that the rebus picture be named, recall of the sentence was not required; subjects simply decided whether the sentence was plausible. As before, it was assumed that if rebus pictures must be named (i.e., if an appropriate lexical representation must be retrieved) before the sentence can be understood, decisions about sentences ending with pictures would take about 200 ms longer than decisions about the corresponding all-word sentences. If the pictured concept is sufficient, however, there should be little or no difference between rebus and all-word sentences. In Experiment 3 and subsequent experiments, a computer-controlled display was used instead of film.

Method

Subjects. The 16 subjects were from the pool described previously. None had participated in the earlier experiments.

Materials, design, and apparatus. The stimulus materials consisted of 48 sentences varying in length from 9 to 13 words ($M = 10.6$). The final word named a picturable object. To produce the implausible versions of the sentences, the final words (pictures) were interchanged between sentences. Thus, there were four forms of each sentence: plausible or implausible, ending in a picture or a word. Examples of the sentences are given in the Appendix. Four versions of the materials were constructed, each containing 12 sentences of each of the four types, in random order. There were 8 additional practice sentences. Each group of 4 subjects saw one version of the materials.

The RSVP sentences were presented on a CRT using a TERAK micro-computer. The words were centered on the screen, as were the rebus pictures. The pictures were line drawings similar to those used in the earlier experiments, entered into the TERAK graphics memory using a HiPad digitizer. Pictures were held in a buffer that, like the words, allowed full presentation in a single scan (16.7 ms). The plausibility decision was made by pressing one key for *yes* and another for *no*, with the right and left hands, respectively; RT was measured from the onset of the final word or picture.

Procedure. Each trial was initiated when the subject pressed the space bar. A row of three asterisks appeared for 300 ms, followed by a 200-ms blank interval and the sentence, presented at 10 words per second. The words were in lowercase letters, except that the final word corresponding to the rebus picture was in uppercase letters.

Subjects were asked to repeat aloud the first four practice sentences, after making the plausibility decision, to make sure that they were able to read most or all of the words (one subject was replaced because of marked difficulty in reading the practice sentences). For the remaining practice trials and the rest of the experiment, subjects were encouraged to make their responses as rapidly as possible, on the basis of their first impression of the plausibility of the sentence. Subjects completed the experiment independently, although the experimenter remained in the room.

Results and Discussion

Correct RTs longer than a subject's mean plus 2 *SDs* were truncated to that number (4.4% of rebus sentences, 4.7% of all-word sentences). Mean RTs and error rates are shown in Table 3. Analyses of the correct RTs showed that responses to all-word sentences were marginally faster (by 22 ms) than responses to rebus sentences, $F_1(1, 15) = 4.54, p = .05, F_2(1, 47) = 1.25, p > .25$. Plausible sentences were judged 70 ms faster than implausible ones, $F'_{\min}(1, 61) = 7.64, p < .01$. There was no significant interaction between modality and plausibility, $F_1(1, 15) < 1.0; F_2(1, 47) = 2.38$. As Table 3 shows, however, the marginal word advantage was confined to the implausible sentences. Similar analyses of the errors showed no significant effects; all *Fs* were less than 1.0. Inspection of Table 3 shows that the error rate was low in all conditions. The results confirm the main finding of Experiments 1 and 2: Rebus sentences are not substantially more difficult to read and understand than are all-word sentences. That is true even when the critical picture appears at the end of the sentence and hence there is no extra time to retrieve the lexical entry in parallel with reading the sentence (as there may have been in Experiments 1 and 2). The statistically marginal 22-ms word advantage is an order of magnitude smaller than would be expected if subjects did have to retrieve a lexical

representation for the picture before they could assess the plausibility of the sentence. Thus, the results conflict with the view that lexical representations are essential in sentence processing.

Experiment 4

A marked difference in naming latency between words and pictures has been obtained when stimuli are presented in isolation, as in the pretest for Experiment 1 (see Table 1). Because it is this difference that is crucial in the logic underlying the present experiments, it is important to show that the difference also holds for words and pictures presented in a sentence context. Suppose it should turn out that picture-naming latency is close to word-naming latency when both are in a sentence context; if so, our claim that rebus pictures are integrated into the sentence without lexical retrieval would be undermined. In Experiment 4, subjects named pictures and words presented as the last item in a sentence, using the materials of Experiment 3. One group of subjects saw the sentences and named the last word or picture (which was plausible or implausible in context). A second group named the same words and pictures following a neutral sentence: "The next item is the. . ." The rate of presentation was 10 words per second, as in Experiment 3.

Method

Subjects. The 24 subjects were from the pool described previously. None had participated in the earlier experiments. Sixteen were in one group, 8 in the other.

Materials and design. The same sentences as those used in Experiment 3 were used for Group 1 ($N = 16$). As in Experiment 3, there were four versions of the experiment, counterbalancing pictures versus words and plausible versus implausible sentence context. A neutral sentence context, "The next item is the," was used for Group 2 ($N = 8$). For this group there were two versions of the experiment, counterbalancing words and pictures. As before, the order of the words and pictures was random. The critical word was capitalized.

Procedure. Except as specified, the procedure was the same as that of Experiment 3. Subjects were instructed to name the last word or picture as rapidly as possible; latency was measured, by a voice key, from the onset of the critical item. The experimenter recorded the subject's response. In Group 1, subjects were told that reading the sentence would help them to respond more rapidly; they were also told that the sentences could be plausible or implausible. In Group 2, subjects were told what the neutral sentence would be. As in Experiment 3, subjects were asked to repeat aloud the first few practice sentences. No subject had unusual difficulty in doing so, so none was excluded from the experiment.

Results and Discussion

Correct RTs longer than a subject's mean plus 2 *SDs* were truncated to that number (in Group 1, 3% of sentences with pictures and 2% of all-word sentences; in Group 2, 2% of pictures, 3% of words). The mean RTs and error rates are shown in Table 3 (synonyms or semantically close names for the pictures—about 9% of the trials—were accepted as correct). Because error rates were low, no further analysis of the errors was carried out. In addition to naming errors, 4% of the responses in Group 1 and 5% in Group 2 were omitted from the

Table 3

Mean Reaction Times (in Milliseconds), Standard Deviations, and Error Rates in Judging the Plausibility of Rebus and All-Word Sentences (Experiment 3) and in Naming the Pictures and Words in Those Sentences or Neutral Sentences (Experiment 4)

Sentence type	Rebus			Words			Word advantage	
	RT	SD	E	RT	SD	E	RT	E
Experiment 3: Plausibility judgment								
Plausible	794	194	.06	788	172	.05	6	.01
Implausible	880	204	.05	842	202	.04	38	.01
Experiment 4: Naming								
Group 1								
Plausible	990	129	.00	806	81	.01	184	-.01
Implausible	1,120	167	.04	807	73	.01	313	.03
Group 2								
Neutral	1,016	112	.04	794	83	.01	222	.03

RT analyses because the subject made an irrelevant sound before responding, or the voice key failed to respond.

Analyses of variance on subject means and item means were carried out. For Group 1, who named pictures and words in plausible and implausible sentence contexts, the 248-ms advantage of words was significant, $F'_{\min}(1, 40) = 45.3$, $p < .001$, as was the 65-ms advantage of plausible over implausible sentences, $F'_{\min}(1, 45) = 7.25$, $p < .01$. The interaction of these two factors was also significant, $F'_{\min}(1, 46) = 6.81$, $p < .05$. The interpretation of this interaction is simple: Whereas there was a large plausibility effect for pictures, there was none at all for words. Even for plausible sentences, however, the 184 ms faster response to words than to pictures was highly significant ($p < .001$).

In Group 2, with a neutral sentence context, words were named 222 ms faster than pictures, $F'_{\min}(1, 17) = 37.9$, $p < .001$. In an analysis comparing Group 1's responses to plausible sentences with Group 2's responses to neutral sentences, there was no significant overall difference (both F_1 and F_2 were less than 1.0) and there was no interaction between group and the modality of the stimulus in the subjects analysis ($F_1 < 1.0$), although $F_2(1, 47) = 4.71$, $p < .05$. The main effect of modality was highly significant, $F'_{\min}(1, 58) = 50.5$, $p < .001$. In a comparison of Group 1's responses to implausible sentences with Group 2's responses to neutral sentences, Group 1's responses were somewhat slower, overall, although the difference was significant only in the items analysis, $F_1(1, 22) = 1.83$, $p < .20$, $F_2(1, 47) = 10.4$, $p < .01$. There was a suggestion of an interaction with modality, $F_1(1, 22) = 2.43$, $p < .14$, $F_2(1, 47) = 5.7$, $p < .03$. Inspection of Table 3 indicates that words were virtually unaffected by an implausible context, whereas pictures were 104 ms slower when the context was implausible rather than neutral. Again, the main effect of modality was highly significant, $F'_{\min}(1, 57) = 56.4$, $p < .001$.

In comparing the overall latencies with those of the naming pretest for Experiment 1 (Table 1), it is evident that naming a word or picture at the end of a sentence (even a neutral sentence) is slower than naming with no context. The stimuli and apparatus were different in the two experiments, so the compar-

ison must be interpreted cautiously, but the result suggests (not surprisingly) that processing the context sentence takes some capacity. (Note, however, that the size of the picture-word difference was at least as great with a sentence context as with no context whatever.)

The main result from both groups in Experiment 4 is that pictures take markedly longer to name than words when presented in sentences, just as they do when presented in isolation. This finding confirms the assumption that a lexical representation for a picture is available later than that for a word even in a sentence context. A second result is also important: The appropriateness of the sentence context had no effect on word naming, but had a dramatic effect on picture naming. This suggests that the lexical entry is contacted independently of sentence context in the case of words (Forster, 1981), but only subsequent to the analysis of the whole sentence (including the pictured concept) in the case of pictures. This result provides strong support for the conceptual view of sentence processing.

In pilot work for Experiment 4, a group of 8 subjects was instructed to name the final item and then report the plausibility of the sentence. With this instruction, responses to words were slowed by 206 ms and responses to pictures by 89 ms, reducing the picture-word difference to 132 ms, and a substantial plausibility effect was obtained for words (a 47-ms effect) as well as pictures (a 111-ms effect). This suggests that subjects can name words either at the point of initial lexical access or can delay their response to a later point, subsequent to sentence processing. This may account for inconsistent results in the literature concerning the influence of sentence context on naming latency.

Just what is meant by retrieval of a lexical representation, and how is retrieval related to overt naming? An assumption we have made is that differences in naming latency for pictured objects and written words reflect differences in the time required to access the mental representation of that word. Notice that because the overt response is the same whether the stimulus is a picture or written word, motor planning and output can be disregarded. Because no parallel difference between pictures and words is found in tasks requiring understanding without

naming (such as matching to a category or deciding on the plausibility of a sentence), the large difference in naming latency cannot be attributed to different speeds of perceptual recognition. A plausible explanation for the extra time required to name pictures is that a picture's concept is retrieved first, and a lexical entry for that concept is retrieved only subsequently (e.g., Potter & Faulconer, 1975; Snodgrass, 1984).

More specifically, we have assumed the relative time to retrieve one part of a lexical entry—information about pronunciation—is a good measure of relative time to retrieve other aspects of the entry, including whatever semantic information is part of that entry. That is, the entry is assumed to become available as a whole. No doubt this is an oversimplification. The components of a word's representation (its orthography, phonology, articulation, and syntax, leaving aside the present question about semantics) are separable under some circumstances, as work in neurolinguistics makes clear (e.g., Allport & Funnell, 1981). Nonetheless, it is a safe assumption that in normal subjects there are close links among these aspects of a word's representation. Forster (1981) argues, for example, that naming latency is a pure measure of relative lexical retrieval time for written words. Neither in Forster's work nor in this study is it claimed that the production of a phonological code is *required* during reading, however; that is a separate question. Naming latency is simply used as an index of relative time to retrieve the whole lexical entry, including whatever semantic information is in the lexicon (rather than the conceptual system).

Experiment 5

Composition (combination) of word meanings is necessary to understand a sentence. The lexical approach, we have suggested, assumes that composition is based on syntactic and semantic information provided in the lexicon. That assumption is undermined by the evidence from Experiments 1–4 that rebus pictures can enter into the composition of sentence meaning without contacting the corresponding lexical representation. Instead, semantic composition would appear to take place in an amodal conceptual system; words (like pictures) merely serve as pointers to the relevant part of the conceptual system.

A more focused test of the locus of semantic composition was made in Experiment 5, in which pictured objects and equivalent nouns were presented with and without a preceding adjective. An adjective and noun in a noun phrase constitute the paradigm case of semantic composition. A concept such as *old chair* is not simply the notion of *old* plus the notion of *chair*, but some notion more like *chair that is old for a chair*. The putting together of such complex ideas might be guided by information in the lexicon, if the lexical approach is right. In that case, a rebus picture should not combine as readily with a preceding adjective as would the corresponding noun, even under conditions when pictures without adjectives are readily understood.

Method

Subjects. The 24 subjects were from the same pool used in previous experiments; none had participated in any of the previous experiments.

Materials and apparatus. The materials for the experiment consisted

of forty-eight 12-word sentences (11 words, when the adjective was omitted), each of which included a concrete noun that was replaced by a picture in the rebus version of the sentence. The noun or picture could be preceded by an adjective; the adjective was selected so that the plausibility of the sentence was not appreciably affected by its omission. Furthermore, the adjective was neutral with respect to the picture: The picture was an appropriate illustration of the equivalent phrase whether the adjective was included or deleted. Examples are *rented house* and *ripe strawberry*. Half the sentences (selected randomly) were made implausible by replacing a plausible final word with an implausible one, (e.g., "A *ripe strawberry* dropped from the basket and rolled across the sky"). Other examples are given in the Appendix.

Eight plausible filler sentences (which formed a separate pilot experiment) were intermixed with the 48 main sentences. The adjectives in two of these sentences in each version were chosen to be incompatible with the rebus picture; examples are *sliced lemon* (with a picture of a whole lemon) and *uprooted tree* (with an upright tree). The results of this pilot were generally consistent with the main experiment and will not be discussed further.

The RSVP sentences were presented on a CRT using the method described in Experiment 3. The plausibility decision was made by pressing one key for *yes* (with the right hand) and another for *no* (with the left hand). Recall was spoken.

Design and procedure. Four versions of the materials were prepared in which a given sentence appeared with or without the critical adjective, and with a rebus picture or all in words. Half the 48 sentences were implausible; this factor was not varied for a given sentence. In each version, there were equal numbers of sentences with and without adjectives, with and without pictures, and plausible versus implausible. (The 8 filler sentences were all plausible, however, so overall there were 32 plausible sentences and 24 implausible sentences in each version.) There were 7 additional practice sentences.

Each trial was initiated when the subject pressed the space bar. A row of three asterisks appeared for 300 ms, followed by a 200-ms blank interval and the sentence, presented at 10 words per second. Except for the first letter of the first word, the words were in lowercase letters. Subjects were instructed to decide on the plausibility of the sentence and then recall it aloud. The experimenter recorded any errors in recall.

Results and Discussion

To preview the main results, rebus sentences without adjectives were as rapidly and accurately comprehended and about as accurately recalled as the corresponding all-word sentences. Rebus sentences with adjectives, although as accurately comprehended as the all-word controls, were less accurately recalled, and there was also some suggestion that those sentences were accepted as plausible more slowly than sentences in the other conditions. As will be seen, however, this rebus-adjective impairment had an unanticipated pattern: The effect was restricted to familiar noun phrases. This pattern, we will argue, is not what would be expected if lexical entries are needed for noun phrase composition.

Plausibility judgments. The results of the plausibility judgment are shown in Table 4. The overall error rate was .11; no analysis of errors was carried out because inspection showed that the error rate did not vary substantially with any of the variables. In particular, there was no difference in error rates between rebus and all-word sentences or between sentences with the adjective and those without, nor was there an interaction; all four means were .11.

Analyses were carried out on correct RTs after replacing long

Table 4

Experiment 5: Reaction Times (in Milliseconds), Standard Deviations, and Error Rates in Judging the Plausibility of Sentences With and Without an Adjective

Noun phrase	Plausible						Implausible					
	Rebus			Words			Rebus			Words		
	RT	SD	E	RT	SD	E	RT	SD	E	RT	SD	E
Adjective	1,146	354	.09	1,121	325	.11	1,216	250	.13	1,226	300	.12
No adjective	1,096	337	.08	1,123	359	.14	1,237	275	.14	1,229	331	.08

RTs with a given subject's mean plus 2 *SDs* (4.9% of the trials). Only sentence plausibility had a significant effect, with plausible sentences classified 106 ms faster than implausible sentences, $F_1(1, 20) = 7.5, p < .02, F_2(1, 46) = 8.14, p < .01$. All other *F*s, for both main effects and interactions, were less than 1.10. Nonetheless, inspection of the means shows that, for plausible sentences, the presence of an adjective increased the RT to rebus sentences by 50 ms, but did not affect all-word sentences. This pattern, which is what would be expected if the lexical approach is correct, was not present in implausible sentences, but that could mean that the subjects were not obliged to process the rebus adjective in order to confirm the implausibility of the sentence. Thus, even though these effects were not statistically significant, they were sufficiently like those predicted from the lexical approach to make one cautious about accepting the null hypothesis. (However, see the discussion below for further illumination of this result.)

Recall. Three recall measures were tabulated: (a) recall of the noun or picture, (b) recall of the adjective, and (c) recall of the other 10 words of the sentence. Although recall accuracy was expected to be sensitive to processing difficulty, it was not obvious in which of the three measures difficulties would be manifest. As it turned out, only measure (c) (recall of the other 10 words) showed significant effects of the experimental variables. On measure (a), the critical noun was recalled incorrectly in .11 of the sentences with an adjective and .10 without; the corresponding error rates for the rebus picture were .12 and .08 (*ns*). On measure (b), given correct recall of the critical picture or noun, the adjective was omitted in .31 of the rebus sentences and .41 of the all-word sentences (not significant, by a binomial test).

Concerning recall measure (c), the proportions of recall errors among the remaining 10 words in the sentence are shown in Table 5. Recall was marginally more accurate for sentences without adjectives than for those with adjectives (.16 vs. .19 of the words were omitted, respectively), $F_1(1, 23) = 2.58, p < .13, F_2(1, 46) = 9.92, p < .01$. Recall was also marginally more accurate for all-word than for rebus sentences, (.16 vs. .18 omissions), $F_1(1, 23) = 3.67, p < .07, F_2(1, 46) = 5.61, p < .03$. Plausible sentences were better remembered than implausible sentences (.14 vs. .20 omissions) in the subjects analysis, $F_1(1, 20) = 29.05, p < .001$, but not in the items analysis, $F_2(1, 46) = 2.67, p > .10$ (note that plausibility was a between-sentences variable).

The main effects were qualified by second-order interactions between rebus format and each of the other variables. Rebus/

word format and plausibility interacted, $F_1(1, 23) = 5.49, p < .03, F_2(1, 46) = 5.55, p < .03$. A Newman-Keuls test showed that implausible rebus sentences were less accurately recalled (.22 omissions) than plausible rebus sentences (.17 omissions), which were less accurately recalled than either plausible or implausible all-word sentences (.14 omissions). More to the point, there was an interaction between format and the presence or absence of an adjective, $F_1(1, 23) = 6.23, p < .02, F_2(1, 46) = 5.40, p < .03$. A Newman-Keuls test showed that rebus sentences with adjectives were recalled less accurately (.21 omissions) than other types of sentences (.16 omissions), which did not differ. The triple interaction of format, plausibility, and adjective condition was not significant; F_1 and F_2 were both less than 1.

Lexical composition versus lexical lookup of stock phrases. The observed increase in the difficulty of recalling a rebus sentence when it includes a picture preceded by an adjective seems to support the claim that composition of word meanings depends on semantic information in the lexicon, information that would be less readily available for a picture. There is, however, another possible explanation of the result: At least some of the adjective-noun phrases (such as *seedless grapes*) could have been familiar to subjects and therefore simply recognized as whole phrases. The integrity of the phrase would be destroyed by replacing the noun with a picture, so that lexical lookup of the phrase as a whole could not occur. It has been noted by many authors (e.g., Lyons, 1977) that, with frequent use, certain combinations of words become sufficiently familiar that they become much like a true compound expression. The meaning of such familiar, partially frozen phrases may not be

Table 5

Experiment 5: Proportions of Words Omitted and Standard Deviations in Recall of Rebus and All-Word Sentences With and Without an Adjective

Noun phrase	Plausible				Implausible			
	Rebus		Words		Rebus		Words	
	P	SD	P	SD	P	SD	P	SD
Adjective	.17	.097	.15	.081	.26	.170	.18	.089
No adjective	.12	.080	.14	.088	.19	.134	.17	.124

Note. Errors in recall of the 10 words other than the adjective (if any) and the critical noun or picture. P = proportion of words omitted.

computed from the meanings of the individual words, but instead may be retrieved as a whole (e.g., Potter & Faulconer, 1979; Swinney & Cutler, 1979). That is, the phrase may act as a single lexical entry, *seedless grapes*, like outright compounds such as *hot dog*.

Although the phrase *seedless grapes* (unlike *hot dog*) can be understood by the normal compositional process, the compound route might be faster because it would require retrieval of one lexical entry, not two, and semantic composition would be bypassed entirely. Substituting a picture for the noun would prevent use of the compound route (unless the picture's name was retrieved first): Only the compositional route would be available. Therefore, for a stock phrase that might be recognized as a single lexical unit, having both the adjective and noun as words should make for easier processing than having the adjective plus a picture.

Note the difference between the stock-phrase hypothesis and the lexical versus conceptual hypotheses: The stock-phrase hypothesis concerns which strings of letters count as a single lexical item or word (*Hot dog* does; what about *seedless grapes*?), and this question is separate from the issue of where word meanings are represented. Whether *seedless grapes* has a single lexical entry, one entry for each of the two words, or (more plausibly) all three entries, the meaning of each of these entries could be represented either in the lexicon itself or in the conceptual system. The problem for the present research strategy, however, is that stock phrases would give an advantage for words over pictures that might be mistaken for a lexical advantage in semantic composition, even though the point about stock phrases is that they do not require composition because they already exist as single lexical entities. Fortunately, there is a way to determine post hoc whether the (small) word advantage observed in Experiment 6 is attributable to superior semantic composition (the lexical hypothesis) or to stock phrases. If the stock-phrase hypothesis is correct, then only the more familiar phrases should show a word (vs. picture) advantage. If, however, the observed word advantage means that semantic composition is better for two words than for a word plus a picture, that result should be at least as evident for unfamiliar phrases as for familiar phrases.

A post hoc test of the stock-phrase hypothesis was carried out, using ratings of phrase familiarity obtained from eight new subjects. Nineteen phrases out of the 48 were rated as definitely familiar (e.g., *seedless grapes* and *copper kettle*). A breakdown of the results into those 19 sentences and the remaining 29 sentences supported the stock-phrase hypothesis. The familiar phrases showed a .08 word advantage in recall (measure [c]), whereas the unfamiliar phrases showed only a .02 word advantage. (The RT results showed the same general pattern.)

Conclusions. A determined effort was made to find some measure of performance in Experiment 5 that would show an advantage for all-word sentences over rebus sentences. The only significant word advantage, which appeared in just one measure (recall of the other words of a sentence), seems to have been due to a different factor entirely, namely, the compound-like processing of familiar adjective-noun phrases. For the less familiar phrases, in which composition of meaning would be necessary, there was no advantage for all-word sentences over rebus sentences. In the sentences without adjectives, there was no

difference between rebus and all-word sentences in response times (1,167 and 1,176 ms, respectively), error rates in judging plausibility (.111 and .108, respectively), or recall errors (.844 and .845, respectively). Overall, then, no support was obtained for the lexical approach.

General Discussion

The general question with which we began was whether aspects of a word's meaning are included in its lexical representation, or whether all semantic information is contained in a non-linguistic conceptual system. These two possibilities give two very different theoretical pictures of the relation between language and thought. The lexical approach has assumed that semantic information in the lexicon is used to arrive at a literal meaning for the sentence; this meaning is then passed to the conceptual system for further nonlinguistic interpretation. In contrast, the conceptual approach places the process of establishing a sentence's meaning in the general-purpose conceptual system.

Because a *pictured object is conceptually but not lexically* equivalent to the corresponding noun, the ability to understand a rebus sentence (in which a picture replaces a noun) would support the hypothesis that a noun acts as a conceptual element. Difficulty with understanding a rebus sentence would support the hypothesis that a noun taps specifically lexical information during the composition of sentence meaning. In the present experiments, sentence processing was pushed close to its temporal limits by using RSVP at rates of 10 or 12 words per second, in order to reveal any differential effects of including a picture.

The main result was that rebus sentences were only marginally more difficult to understand and remember than equivalent all-word sentences. With the sentences used in Experiments 1 and 2, this difference was most evident in the time to make a judgment about sentence plausibility, which took about 104 ms longer on rebus sentences. With two new sets of materials in Experiments 3 and 5, there was no reliable rebus disadvantage except in one circumstance to be discussed shortly. Thus, overall, rebus sentences did not produce a consistent deficit in speed or accuracy of comprehension or in accuracy of immediate recall. It made no difference whether the picture appeared in the middle of the sentence (Experiment 5) or at the end of the sentence (Experiment 3) and nor did it matter whether there was one picture or two (Experiment 1). Moreover, the rebus format did not interact significantly with other variables in these experiments, such as plausibility and sentence length. The absence of interactions suggests that the rebus picture was represented in a sentence-compatible form at an early stage of processing, before these other variables came into play.

Experiment 5 tested the hypothesis that a rebus picture would have special difficulty combining with a preceding adjective. A small word advantage was obtained in one special case: when the phrase was rated as familiar (e.g., *alarm clock*). When such phrases were expressed in words they could evidently be recognized as a single lexical unit and therefore no semantic composition (whether lexical or conceptual) was required. When the noun was a picture, however, the lexical unit was broken up and semantic composition was necessary. When the

phrases were unfamiliar, and both all-word phrases and word-plus-picture phrases required semantic composition, the word advantage disappeared. This indicates that semantic composition (as distinct from lookup) was no harder for phrases with pictures than for all-word phrases. This result supports the conceptual approach.

RSVP and On-Line Processing

The logic of the present experiments was to present sentences so rapidly that a delay in encoding the rebus picture into the needed form (such as by silently naming it) would be highly disruptive. For the experiments in which the picture appeared in the middle of the sentence (Experiments 1, 2, and 5) the argument depends on the assumption that the sentences were processed on line, that is, word by word during presentation. If, instead, sentence-level composition began only after presentation of the whole sentence, the name of the rebus picture might have been retrieved by the time it was needed in processing. This problem did not arise in Experiment 3 because the critical picture or word was at the end of the sentence, and yet there was only a marginal word advantage in that experiment. Still, it is worth reviewing briefly the evidence that processing of an RSVP sentence does indeed take place on line.

Four kinds of evidence support on-line processing:

1. If readers simply stored the words as a list during presentation and processed the sentence afterward, decision errors or postsentence time to make the plausibility decision or both would be expected to increase with the length of the sentence. But no such systematic increases were observed (Experiments 1 and 2).

2. Unrelated words are difficult to retain when presented at 10 or 12 words per second, even when the list is as short as 5 words (Potter, 1982), whereas the 8- to 14-word sentences used in the present experiments were recalled fairly accurately.

3. Although the order of unrelated items is known to be poorly retained when they are presented at rates of six or more items per second (e.g., Scarborough & Sternberg, 1967; Sperling & Reeves, 1980), order errors in recall were low in the present experiments (under 2%). To the extent that readers incorporate the words of a sentence into an interpreted structure as the words arrive, order will be preserved. Constraints on the possible word order of an unordered set of words might also contribute to the accuracy of recall, but subjects are not very good at reconstructing the order of scrambled RSVP sentences. For example, in an experiment using the materials of Experiment 1 (Potter & Kroll, 1984), 26% of the words recalled were misordered, compared with 1% in a comparison group with normally ordered RSVP sentences.

4. The pattern of errors in immediate recall also supports on-line processing. Typically, the words and phrases relevant to the main topic of the sentence were the ones omitted in recall. That result can be explained in several ways, but all the explanations assume that the subject understood at least the gist of the sentence during presentation (e.g., Frauenfelder, Dommergues, Mehler, & Segui, 1979). A more detailed discussion of this and other evidence for on-line processing is provided in Potter (1984; see also Masson, in press).

RSVP was used in the present experiments to control the

time available for possible compensatory processing of the rebus picture. An experiment in which the sentences of Experiment 1 were presented in a conventional simultaneous format gave much the same results, however (Potter & Kroll, 1984).¹ This reinforces the conclusion of Potter (1984) that RSVP reading, although speeded, draws on the same processing mechanisms as conventional reading.

When a Name Is Required: Rebus Puns and German Gender

In Experiment 4 subjects named words and pictures that appeared in appropriate, neutral, or inappropriate sentence contexts. Latency to name pictures was more than 200 ms longer than to name words, confirming earlier results for pictures and words in isolation, and showing that a lexical entry for a picture (even in a sentence) is not available as soon as that for a written word. Using a method that did not require overt naming, Potter (1981; see Footnote 1) had subjects judge the plausibility of RSVP sentences in which the rebus picture's name was a homonym such as *lock*. In half of the rebus sentences the picture illustrated the wrong meaning of the word, as in the following, in which a line drawing of a padlock replaced the word *lock*: "The girl curled a *lock* of her hair nervously around her finger." If readers routinely understand rebus sentences by covertly naming the picture (perhaps at an abstract morphophonemic level, Taft, 1984), such a picture pun might present no special difficulty. One group of subjects was instructed to decode such picture puns and accept the resulting sentence as plausible when the corresponding word would have been plausible. They were slower and made more errors than another group of subjects instructed to respond in terms of the pictured object itself (and hence to reject the picture puns). Thus, forcing the use of a purely lexical representation of a rebus picture created difficulties, whereas forcing the use of a purely conceptual representation did not.

In a second experiment (Potter & Carpenter, 1984, see Footnote 1) a given subject saw only two sentences with picture puns, intermixed with a larger number of other rebus and all-word sentences. Without any advance warning that such picture puns might occur, subjects spontaneously and rapidly rejected the pun 71% of the time. On the other 29% of the trials the pun was accepted, but responses were significantly slower than for other rebus sentences. These experiments provide evidence against a strong version of the lexical approach in which a rebus picture is replaced during sentence processing by a single lexical representation, one that is ambiguous in the case of homonyms such as *lock*. (See Baron, 1973, for experiments using an analogous design to reject phonological representations as the sole mediating representations used in reading.)

On the question of whether gender agreement requires a retrieval of a lexical entry, Schönle, Schön, and Friedhoff (1982) showed German subjects pictures or written words accompanied by a masculine, feminine, or neuter article (*der*, *die*, or *das*). The task was to decide as rapidly as possible whether or

¹ Details of Potter and Kroll (1984), Potter and Carpenter (1984), and Potter (1981) are available from the first author.

not the article was correct (as noted earlier, a noun's gender in German is frequently unrelated to the masculine, feminine, or null sex of the referent). For correct yes responses, the mean RT was 360 ms longer for pictures than for words. Once again, when specifically lexical information was required, pictures took substantially longer than words.

Pictured Versus Named Concepts

Pictures successfully replaced concrete nouns in the present experiments. Could pictures replace words other than nouns, such as verbs or adjectives? We consider that unlikely, because any picture of an action is also a picture of one or more objects (actors) carrying out the action, and almost any picture of a property also depicts an object. Thus, pictures are ill suited to depict verbs and adjectives that are separate from nouns. The point of the present experiments, however, was not to show that pictures can replace words in general (which they obviously cannot), nor to support an imagery-based theory of meaning, but to make use of the already established equivalence of some pictured objects and their (basic level) names to address a specific issue, the representation of word meaning. We have shown that there is one category of words (concrete nouns) whose meaning enters into sentence processing at a conceptual level rather than a purely lexical level. That makes it unlikely that the meanings of other categories of content words enter at a different level (viz., lexical), for composition of meaning could not take place unless all semantic information were represented at the same processing level.

On Modularity

Do the present results conflict with the hypothesis that language perception is modular (J. A. Fodor, 1983)? It depends on what the modules are taken to be. Fodor believes that the language module (which is itself a loosely interconnected set of processing units) has limited interpretive functions. All the language processor does is "deliver, for each input utterance, a representation which specifies its lexical constituents" (i.e., the lexical type that each token word belongs to; p. 91) and probably a syntactic analysis of the sentence at the level of logical form. Fodor acknowledges that even this level of output may require considerable computation (see, for example, 1983, Note 28, p. 135), and he is therefore noncommittal about just where to draw the line between the language module and nonmodular cognitive processes. It would seem clear, however, that an automatic syntactic analysis that is entirely internal to the language module could not be carried out on a sentence in which a picture substitutes for a noun, without first identifying an appropriate lexical type for the picture—and we have argued that that does not happen.

The present study has said nothing directly about syntax, except to suggest that sentences must have been parsed successfully or they would not have been understood well enough to evaluate their plausibility and to recall them when presented at 10 or 12 words per second. Because rebus pictures could be processed successfully in such sentences, the syntactic processor is unlikely to be fully modular: The nounlike character of a pictured object must be understood by the parser.

Conclusion

The purpose of the present research was to evaluate two approaches to the representation of word meaning. Are words in sentences understood and their meanings combined by virtue of semantic information represented in the lexicon? Or, does a word's lexical representation simply point to a general-purpose conceptual system in which the meanings of words are retrieved and combined? The finding that a pictured object can successfully replace a noun in a rapidly presented sentence supports the second of these approaches.

References

- Aaronson, D. (1976). Performance theories for sentence coding: Some qualitative observations. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 42–55.
- Akmajian, A., Demers, R. A., & Harnish, R. M. (1979). *Linguistics: An introduction to language and communication*. Cambridge, MA: MIT Press.
- Allport, D. A., & Funnell, E. (1981). Components of the mental lexicon. *Philosophical Transactions of the Royal Society of London*, B 295, 397–410.
- Banks, W. P., & Flora, J. (1977). Semantic and perceptual processes in symbolic comparisons. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 278–290.
- Baron, J. (1973). Phonemic stage not necessary for reading. *Quarterly Journal of Experimental Psychology*, 25, 241–246.
- Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, 11, 63–65.
- Clark, H. H., & Clark, E. V. (1977). *Psychology and language: An introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich.
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Fodor, J. D. (1977). *Semantics: Theories of meaning in generative grammar*. New York: Crowell.
- Forster, K. I. (1970). Visual perception of rapidly presented word sequences of varying complexity. *Perception & Psychophysics*, 8, 215–221.
- Forster, K. I. (1981). Priming and the effects of sentence and lexical contents on naming time: Evidence for autonomous lexical processing. *Quarterly Journal of Experimental Psychology*, 33A, 465–495.
- Fraisse, P. (1960). Recognition time measured by verbal reaction to figures and words. *Perceptual & Motor Skills*, 11, 204.
- Frauenfelder, U., Dommegues, J. Y., Mehler, J., & Segui, J. (1979). L'integration perceptive des phrases [The perceptual integration of sentences]. *Bulletin de Psychologie*, 32, 893–902.
- Frederiksen, J. R., & Kroll, J. F. (1976). Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 361–379.
- Green, D. W. (1977). The immediate processing of sentences. *Quarterly Journal of Experimental Psychology*, 29, 135–146.
- Levitt, W. J. M., & Kempen, G. (1975). Semantic and syntactic aspects of remembering sentences: A review of some recent continental research. In A. Kennedy & A. Wilkes (Eds.), *Studies in long term memory* (pp. 201–216). New York: Wiley.
- Lyons, J. (1977). *Semantics* (Vol. 2). Cambridge, England: Cambridge University Press.
- Masson, M. E. J. (in press). Comprehension of rapidly presented sentences: The mind is quicker than the eye. *Journal of Memory and Language*.
- Mitchell, D. C. (1979). The locus of the experimental effects in the rapid serial visual presentation (RSVP) task. *Perception & Psychophysics*, 25, 143–149.

- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.
- Paivio, A. (1978). Dual coding: Theoretical issues and empirical evidence. In J. M. Scandura & C. J. Brainerd (Eds.), *Structure/process models of complex human behavior* (pp. 527-549). Leiden, the Netherlands: Nordhoff.
- Potter, M. C. (1979). Mundane symbolism: The relations among objects, names and ideas. In N. R. Smith & M. B. Franklin (Eds.), *Symbolic functioning in childhood* (pp. 41-65). Hillsdale, NJ: Erlbaum.
- Potter, M. C. (1981, November). *Reading picture puns*. Paper presented at the meeting of the Psychonomic Society, Philadelphia, PA.
- Potter, M. C. (1982, November). *Very short-term memory: In one eye and out the other*. Paper presented at the meeting of the Psychonomic Society, Minneapolis, MN.
- Potter, M. C. (1984). Rapid serial visual presentation (RSVP): A method for studying language processing. In D. Kieras & M. Just (Eds.), *New methods in reading comprehension research* (pp. 91-118). Hillsdale, NJ: Erlbaum.
- Potter, M. C., & Carpenter, E. (1984). [The plausibility of sentences with unexpected picture puns]. Unpublished raw data.
- Potter, M. C., & Faulconer, B. A. (1975). Time to understand pictures and words. *Nature*, 253, 437-438.
- Potter, M. C., & Faulconer, B. A. (1979). Understanding noun phrases. *Journal of Verbal Learning and Verbal Behavior*, 18, 509-521.
- Potter, M. C., & Kroll, J. F. (1984). *Sentence comprehension in RSVP*. Unpublished manuscript.
- Potter, M. C., So, K.-F., Von Eckardt, B., & Feldman, L. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior*, 23, 23-38.
- Potter, M. C., Valian, V. V., & Faulconer, B. A. (1977). Representation of a sentence and its pragmatic implications: Verbal, imagistic or abstract? *Journal of Verbal Learning and Verbal Behavior*, 16, 1-12.
- Scarborough, D. L., & Sternberg, S. (1967). *Processing items and their order in sequential visual displays*. Paper presented at the meeting of the Eastern Psychological Association, Boston, MA.
- Schönle, P. W., Schön, G., & Friedhoff, W. (1982). *The distinctiveness of gender representation in the mental lexicon*. Unpublished manuscript.
- Snodgrass, J. G. (1980). Towards a model for picture-word processing. In P. A. Kollers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language* (Vol. 2, pp. 565-584). New York: Plenum.
- Snodgrass, J. G. (1984). Concepts and their surface representations. *Journal of Verbal Learning and Verbal Behavior*, 23, 3-22.
- Sperling, G., & Reeves, A. (1980). Measuring the reaction time of a shift of visual attention. In R. Nickerson (Ed.), *Attention and performance VIII* (pp. 347-360). Hillsdale, NJ: Erlbaum.
- Swinney, D. A., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, 18, 523-534.
- Taft, M. (1984). Evidence for an abstract lexical representation of word structure. *Memory & Cognition*, 12, 264-269.
- Theios, J., & Freedman, K. (1984, November). *The widely reported "advantage" of pictures over words is due to the widespread use of large pictures and small words*. Paper presented at the meeting of the Psychonomic Society, San Antonio, TX.
- Winer, B. J. (1962). *Statistical principles in experimental design*. New York: McGraw-Hill.

Appendix

In these examples of the materials in the experiments, the capitalized word became a picture in the rebus version. For sentences with an implausible version, the implausible last word is shown in parentheses.

Experiments 1 and 2

The fox chased the CHICKEN around the yard. (kitchen)
 The OWL sat in the tree looking around for prey. (women)
 When he left he locked the DOOR and walked up the road. (wall)
 In the story of Snow White the red APPLE that she ate was poisonous. (nutritious)

Double-picture sentences:

Plausible: I left by COAT lying on the BENCH and it got wrinkled.
 Implausible: The refrigerator and COUCH were carried down the STAIRS by the baby.

Experiments 3 and 4

Paul came to work soaking wet because he forgot his UMBRELLA. (ANCHOR)
 The old radio worked after she replaced the PLUG. (SHIRT)

Tom walked up to the donkey and held out a CARROT. (TENT)
 She trimmed off the stem of the rose with the SCISSORS. (CAR)

Experiment 5

Plausible sentences (the adjective is shown in parentheses):
 The (alarm) CLOCK rang loudly so he hit it with his hand.
 The spaniel puppy proudly brought the (brown) SHOE up to the owner.
 She was able to identify the (poisonous) MUSHROOM by its musty odor.
 Implausible sentences:
 The man grabbed the woman's (leather) PURSE and ran into the waves.
 For dessert Ann served several cheeses and (seedless) GRAPES on a glove.
 The (borrowed) TROWEL lay unused on a neglected shelf in the refrigerator.

Received November 18, 1985
 Revision received March 10, 1986 ■