Journal of Experimental Psychology:
Learning, Memory, and Cognition
1989, Vol. 15, No. 6, 1134–1146

0278-7393/89/$00.75

Inferences About Contextually Defined Categories

Gail McKoon and Roger Ratcliff
Northwestern University

The categories referred to in textual descriptions are often defined by their context (Roth & Shoben, 1983). The results of experiments in this article suggest that inferences about the most typical exemplars of such categories are encoded to a high degree during reading. Subjects were presented with sentences that included contextual definitions of categories but did not mention any exemplars of the categories. When the most typical exemplars were presented for later recognition, subjects had difficulty (relative to a control) in correctly responding that the exemplar had not been explicitly presented. But when less typical exemplars were presented for recognition, subjects had no difficulty in rejecting them. The results for the less typical exemplars suggest that inferences about contextually defined categories are not made up of the defining and characteristic features of the categories because the less typical exemplars would match such inferences. Instead, the results suggest that the content of the inference is made up of information relating properties of the most typical exemplar to textual information.

In reading, comprehension processes take a text as input and produce a representation of that text in memory as output. Past research has investigated many different kinds of comprehension processes, including coherence operations that tie together references to the same object and elaborative inferences that add information that was not explicitly stated in the text. However, one aspect of comprehension that has received relatively little attention concerns the meanings of words and sentences and how contextual information determines the specific meanings that are encoded into the memory representation of a text.

The experiments in this article were designed to investigate the encoded meanings of contextually defined categories. An example of a text that expresses information about a category is shown in Table 1. The category of interest in the predicting text is *fruit*, and the relevant contextual information is that it is a fruit squeezed for juice for breakfast. Thus, the contextually defined category is *fruits you squeeze for juice for breakfast*. The experiments described below show that information about likely exemplars of such categories is inferred as part of the meaning of the text, as was suggested by Roth and Shoben (1983), and the experiments examine characteristics of the information that is represented in the inferences.

To frame this investigation, the experiments are motivated from two different perspectives. First, there has been considerable research on the inference processes that occur during reading, and the current experiments demonstrate that information about high typical exemplars of contextually defined categories is one kind of inference that can be encoded to a high degree into the representation of a text. Second, there has also been considerable research into the structure of categories in semantic memory, and our experiments relate to this research by examining the issue of how such categories function in discourse processing. In the following sections, theoretical frameworks from these two research areas are reviewed.

This research was supported by National Science Foundation (NSF) Grant 85-16350 to Gail McKoon and NSF Grant 85-10361 to Roger Ratcliff. We wish to thank Ed Shoben for comments on an earlier version of this article and for making available the materials from Roth and Shoben (1983). We also wish to thank Bob Crowder, Keith Rayner, and anonymous reviewers for their comments.

Correspondence concerning this article should be addressed to Gail McKoon, Psychology Department, Northwestern University, Evanston, Illinois 60208.

Elaborative Inferences

Investigations of the inference processes that occur during reading have usually been designed to demonstrate that some particular type of information is or is not inferred. Recently, we have proposed an alternative framework for research in this area (McKoon & Ratcliff, 1986, 1989). The principal notion is that inferences are not necessarily encoded into the memory representation of a text in an all-or-none fashion, but instead can be encoded in varying degrees. The degree of encoding can vary from not encoded at all to partially encoded to exemplars encoded. A partially encoded inference is a set of features of meaning that does not completely instantiate the inference. The features represent attributes of meaning; by using the term *feature*, no commitment is meant to feature or exemplar models: Attributes of meaning might equally well be encoded as propositions.

Examples of inferences with different degrees of encoding have been provided in several experiments. These experiments have examined the degree of encoding of inferences about predictable events. For example, a text about an actress falling off a 14th-story roof could lead to the inference that the actress died. The data show that this kind of inference is only partially encoded into the mental representation of a text (McKoon, 1988; McKoon & Ratcliff, 1986, 1989; Potts, Keenan, & Golding, 1988).

These experiments determined the degree of encoding of an inference by measuring the degree of match between a test word expressing the inference and the representation of the text in memory. The degree of match was measured by testing

Table 1
An Example of the Materials Used in Experiment 1

Predicting sentence
The young attorney wanted to make sure she had fresh juice for breakfast, so she bought and squeezed the fruit herself.
Control sentence
When the waiter brought breakfast, the attorney squeezed the fruit to make sure it was juicy and ripe, not hard and green.

	Test condition	
Word	Text prime	Neutral prime
Prime	attorney	ready
Target	orange	orange
Prime	attorney	ready
Target	grapefruit	grapefruit

recognition of words. Table 2 shows the procedure with an example study–test list from McKoon and Ratcliff (1986, Experiment 4). The procedure will be described in detail to facilitate explaining the pattern of data that demonstrates partial encoding.

Each list began with two texts presented for reading on a cathode-ray tube (CRT) screen. Following the texts, a series of test items was presented. Each item contained a prime word and a target word. The prime was used to control the retrieval context of the target word, and the subjects were simply asked to read it. For the target word, the subject's task was to decide whether or not it had appeared in either of the studied texts. Each test item was made up of a warning signal, the prime word, and the target word, and the target word was followed by a signal to respond. The subject was instructed to make a response 300 ms after the signal to respond. The signal was presented 350 ms after the target word, so the subject had a total of 650 ms in which to make a decision. This deadline procedure was used to prevent subjects from engaging at test time in slow strategic processes that would

Table 2
An Example of the Delayed Recognition Procedure

Press space bar to begin
The director and the cameraman were ready to shoot close-ups when suddenly the actress fell from the 14th story.
The old man loved his granddaughter, and she liked to help him with his animals when she visited the farm; she also liked the milk and cookies her grandmother provided.
TEST TEST TEST
++++++++ (warning signal)
farm (prime)
table (target)
****** (response signal)
++++++++
actress
dead
******
++++++++
ready
cow
*****
.
.
.

construct inferences not encoded in the mental representation of the text during its original reading. Preventing such slow constructive processes means that responses can be interpreted as being based on the retrieval of information encoded when the text was studied, through the interaction of the retrieval cues (the prime and target words) and the encoded information.

In the McKoon and Ratcliff (1986) experiment (Experiment 4), target words that represented inferences (e.g., *dead* in Table 2) had not been explicitly stated in a text, and so the correct response was *no*. An inference target word was tested either after study of a text that predicted it (like the first text in Table 2) or after study of a control text that did not predict it. Evidence for inferences appeared as an increased tendency toward incorrect *yes* responses after the predicting text relative to the control text. With the deadline procedure, response times did not vary across experimental conditions, so differences among conditions were observed in error rates.

Results showed that the error rate for the inference target words was higher with the predicting than the control text, but this difference was significant only when the retrieval context was the prime from the text (actress dead) and not when the retrieval context was the neutral prime (McKoon & Ratcliff, 1986, Experiment 4). These results were the basis for the conclusion that the inference must be encoded only partially; strong evidence for the inference appeared only when the prime was a word from the text.

In contrast, an inference that is more fully encoded should have a high error rate even when the target word is presented by itself with only the neutral prime. This pattern of data was found for elaborative inferences that have the support of well-known associations. For a sentence about the water in a stream getting colder and colder on a snowy winter day, there would be associations between *freeze* and the explicitly stated meaning of the sentence. McKoon and Ratcliff (1989) showed that the error rate for *freeze* presented as a target word was high even in the neutral priming condition, indicating that these kinds of inferences are more fully encoded.

Experiment 1 in this article was designed to further examine the partial encoding of inferences. We hypothesized that predicting texts like the one shown in Table 1 would lead to at least some degree of encoding of information that matched the most likely exemplar of the category (orange); perhaps, following suggestions by Roth and Shoben, the encoded information would match the most likely exemplar to quite a high degree. We also hypothesized that other, less likely, exemplars of the category (such as grapefruit) would match the encoded information only partially, if at all. Thus, for a single text, one exemplar would match the inferred information in memory to a high degree, and another exemplar would match to a lesser degree. Background on the relations between exemplars of categories and contextual information is reviewed in the next section.

Contextually Defined Categories

When comprehension processes operate on a text to produce a representation of the text in memory, we assume that the meaning of the text and the concepts in it can be described as a collection of features (or propositions). Some of the features will reflect explicitly stated concepts or propositions, and others will reflect inferred information. It has often been suggested that the meaning of a concept can be affected by its context (cf. Anderson & Ortony, 1975). For example, McKoon and Ratcliff (1988) found that the encoded features of common nouns could vary across contexts. Most pertinent to this article, Roth and Shoben (1983) showed that the relations between categories and their exemplars were affected by context.

In their experiments, Roth and Shoben (1983) presented subjects with pairs of sentences. The first sentence mentioned a concept by a category label, and the second referred to the same concept by using its exemplar label. There were two possible first sentences; one made the exemplar typical of the category, and the other made it atypical. An example of a pair for which the exemplar is typical is *Stacy volunteered to milk the animal whenever she visited the farm. She was very fond of the cow.* Changing the first sentence to *Fran pleaded with her father to let her ride the animal* makes the exemplar *cow* atypical. Roth and Shoben found that the time subjects took to read the second sentence was affected by typicality. Reading time was faster when the exemplar was typical of the contextually defined category than when it was not.

Roth and Shoben (1983) suggested several mechanisms by which context could determine the relations between categories and their exemplars. One possible mechanism would be derived from semantic memory models. Context would alter the criteria for category membership, so that not all exemplars would be acceptable as members of the contextually defined category. This could be done in feature models (e.g., Smith, Shoben, & Rips, 1974) by altering the importance of (or weights on) various features, or in network models (e.g., Anderson, 1976, 1983; Collins & Loftus, 1975) by a change in the activation levels of various propositions. In either case, the content of the instantiated inference would be the features or propositions that define and characterize the contextually defined category. For example, for the category *fruits squeezed for juice for breakfast*, the collection of encoded features might include *round; juicy; pulpy; somewhat sweet; a few inches in diameter; room temperature or cold, but not hot; contains vitamin C; might prevent colds;* and *mothers make kids drink it.*

The idea that the representation of the meaning of a contextually defined category is the collection of features that define and characterize the category carries a strong implication. The implication is that more than one exemplar can match the collection. If the category has a most typical exemplar (like *orange* in the example in Table 1), then that exemplar will match best. Another exemplar might match to a smaller degree. Grapefruit, for example, have most of the features of fruits that can be squeezed for juice for breakfast. Other categories might not have a most typical exemplar; *cats* and *dogs* might be equally typical of the category *pets for children*, and so might equally match the collection of features encoded as part of a sentence that mentioned this category.

The hypothesis that contextually defined category is encoded as the features that describe the category was tested in the experiments in this article against an alternative hypothesis. The alternative is that the content of the encoded inference is based more on information about a specific exemplar and its relation to the text than on general semantic information about the contextually defined category. The propositions or features that make up the inference would have more to do with, for example, an attorney having orange juice, an attorney squeezing an orange, and an attorney wanting a fresh orange, and less to do with the features of fruits that allow them to produce juice. The encoded inference would not so much represent the category *fruits that can be squeezed for juice*, but instead would comprise information about one highly encoded exemplar relative to its context. The inference might include some information about the general category, but most of the information would concern the relations between oranges and the other information in the sentence. Specific ways in which this information might be represented are suggested in the General Discussion section.

Experiment 1 was designed to distinguish between the two hypotheses by measuring the amount of match between the encoded text and an exemplar of the contextually defined category that was not the most typical exemplar. The experiment was designed to test the hypotheses only for categories for which there was a single most typical exemplar, with other less typical exemplars. For the example text in Table 1, there is one highly typical exemplar (orange) and one less typical exemplar (grapefruit). Although grapefruit is not the most typical exemplar, it does share most of the features of such fruits. If the mental representation of the inference is made up of general information about the category *fruits that can be squeezed for juice*, then *grapefruit* should partially match this information. While *orange* would have the highest degree of match to the encoded information, *grapefruit* would still match to some degree. The alternative hypothesis is that the inference is made up of information based on the relations between the specific exemplar *orange* and the text. Then there would be almost no match between the inference and *grapefruit* because there was no information in the inference about a grapefruit and about an attorney having breakfast, and there is little information about the general category.

Experiment 1

The purpose of Experiment 1 was, first, to investigate whether the inferred information about contextually defined categories is so fully encoded that evidence for it appears even in the neutral priming condition, a less than optimal retrieval context. A second purpose of the experiments was to examine whether the content of the inferences is more reflective of general knowledge about contextually defined categories or information about a specific exemplar of the category as it relates to the text. The experiments can described most easily by reference to an example of the texts that were used (see Table 1).

The predicting text was written to "predict" that the fruit being squeezed was an orange, that is, to make *orange* the most likely member of the contextually defined category *fruits you squeeze for juice for breakfast.* The predicting texts were modeled after those used by Roth and Shoben (1983). To test for a high degree of match between the most typical exemplar and the representation of the text in memory, the delayed recognition procedure described above was used in Experiment 1. Subjects read either the predicting or the control version of a text. The control version, shown by example in Table 1, mentioned all of the words semantically associated to the predicted category member, but the text as a whole did not strongly point to that category member. Thus, the control texts allowed separation of the inferences based on the meaning of the text as a whole that were of interest in this experiment from inferences that might be based on individual words. There were two retrieval contexts, as shown in Table 1: The most likely category exemplar was presented as a target test word either with a prime word from its text or with a neutral prime word (the word *ready,* used throughout the experiment). If the degree of encoding of information about the category exemplar is high with the predicting text, then the error rate on the target word should be high for the predicting text relative to the control text, and this should be true in both retrieval contexts. The category exemplar should access the mental representation of a predicting text even when it is presented by itself with only the neutral prime.

The predicting texts were written not only to have a most likely category exemplar but also to have a secondary exemplar that fit the context. In the example, grapefruit is a fruit that can be squeezed for juice, but not the most likely one. The materials for Experiment 1 were designed to maximize the likelihood that the primary exemplar (e.g., orange) would be relatively fully encoded. The secondary exemplar would then provide a means for examining the content of the inference. If the content of the inference was general information about the category *fruits that can be squeezed for juice for breakfast* then *grapefruit* should partially match this information. Because of this match, it would be difficult to respond correctly (*no*) to *grapefruit* after reading the predicting text, so the error rate on *grapefruit* would be higher for the predicting than the control text. The degree to which *grapefruit* matched encoded information would determine whether higher error rates appeared in both of the retrieval contexts or only with the prime from the text. If the degree of encoding was not very high, then evidence for the match of *grapefruit* against the encoded information might appear only with the prime from the text. In contrast, if the inference for the predicting text is made up of information about the most typical exemplar in relation to the text, then there would be almost no match between the inference and *grapefruit.* The error rate on *grapefruit* would not be higher with the predicting than the control texts, in either retrieval context.

Method

Subjects and design. There were two separate groups of subjects, 24 in each group. For one group, the test words for the predicting and control texts were always the primary exemplars, and for the other group, always the secondary exemplars. For each group, there were four experimental conditions formed by crossing text type (predicting or control) with retrieval context (prime from the text or neutral prime). The four conditions were combined in a Latin-square design with groups of subjects (six per group) and sets of materials (eight per set). The subjects participated for credit in an introductory psychology class.

Materials. For each of 32 experimental tests, there were two versions, predicting and control. The predicting version was written to point strongly to one primary category exemplar while having a reasonable secondary exemplar. The categories were generally restrictions on common objects, such as animals that are milked on farms, weapons that stab, insects that flutter and have wings, and so forth. The control version was written so that it would not point to the exemplars but would include all of the words semantically associated with the exemplars. Some of these words were highly associated with the primary exemplar and could, by themselves and independently of the meaning of a sentence, lead to encoding of the primary exemplar, so the control sentences usually contained information that would point away from the primary exemplar. All of the predicting and control sentences are listed in the Appendix. The predicting texts averaged 25.5 words in length and the control texts averaged 25.7 words.

The test words for the experimental texts were the primary and secondary exemplars, a topic character to be used as a prime word (e.g., *attorney*), and two other words to be used as a positive test word and its prime.

There were also 32 filler texts, each one sentence (averaging 20.8 words). For each, there were four words used as two positive test words and their primes. There were also 128 words that did not appear in any text that were used for negative test words.

To check that the primary exemplars were actually more typical of their categories than the secondary exemplars and that the secondary exemplars were compatible with their categories to some degree, ratings were collected from an independent group of 32 subjects (see McKoon & Ratcliff, in press-a). The subjects were asked to rate, on a 1–7 scale with 7 the highest score, how compatible an exemplar was with its predicting or (for different subjects) its control sentence. The average ratings were as follows: predicting, primary exemplar = 6.47; predicting, secondary exemplar = 4.38; control, primary exemplar = 2.91; and control, secondary exemplar = 2.47.

Procedure. Subjects were each tested in one 50-min session. All materials were presented on a CRT screen, and responses were collected on the CRT's keyboard. Presentation of the materials and data collection were controlled by a real-time microcomputer system.

To ensure that subjects responded quickly enough that slow, strategic processing could be ruled out, a response deadline procedure was used (McKoon, 1988; McKoon & Ratcliff, 1986, 1989). Subjects were given practice at responding to a deadline with a lexical decision task. The lexical decision task was used because it allows practice on responding to the deadline without the necessity of prior study and so proceeds quickly. Each item for lexical decision began with a row of +s displayed for 500 ms. Then the +s disappeared, and on the line below, a prime word, either *ready* or some other word, was displayed for 250 ms. When the prime disappeared, a test letter string was displayed on the line below where the prime had been. After 250 ms, a row of asterisks was presented below the test string. Subjects were instructed to respond exactly 300 ms after the asterisks appeared, *yes* (by hitting the ?/ key) if the string was a word, *no* (by hitting the *Z* key) if the string was not a word. After the response, the test string was cleared from the screen, and the response time was displayed. If the response time was faster than 200 ms, the message TOO FAST! was also displayed. If the time was slower than 350 ms, the message TOO SLOW! was also displayed. After 450 ms if the response was not too slow or too fast, or after 500 ms if it was, the screen was cleared, and after a blank screen lasting for 550 ms, the warning signal for the next test item was presented. After every 10th item, there was an instruction to press the space bar when ready to begin the next 10 items. There were 90 test items altogether, and subjects reported feeling comfortable with the deadline procedure after this much practice.

After the lexical decision practice, subjects began the study–test procedure. There were 4 practice trials, and then 16 experimental trials. On each trial, there were 4 texts to study and 12 test items.

A trial began with an instruction to the subject on the CRT screen to press the space bar to initiate the trial. Then the four texts were presented, one at a time. On each trial, there were two experimental texts and two fillers. Each text was presented for 5,500 ms, and followed by a 1,000-ms blank interval. After the fourth text, a row of asterisks was presented for 2,000 ms to signal the beginning of the test list. The test items were presented as in the lexical decision practice, except that the signal to respond (asterisks) was displayed 350 ms after the target word. Feedback about response time (too slow or too fast) was also the same as in the lexical decision practice. Subjects were instructed to give their response exactly 300 ms after the response signal, responding *yes* if the test word had appeared in a studied text, and *no* if it had not. After the 12th test item, the instruction to press the space bar to begin the next trial was displayed. Subjects were told to attend to the primes because they might facilitate responding to the target words.

Two of the target words in the list of 12 on each trial were the category exemplars for the experimental texts. These were primed according to their condition, either by the topic word from their text or by the neutral prime, *ready.* There were also one positive filler target word from each experimental text and two positive and two negative target words for each filler text. The positive and negative fillers were primed either by a word from their text (probability, .7) or the neutral prime (probability, .3). No word was repeated in the test list, except the neutral prime, *ready.* The probability of an item in which a word from a text primed an exemplar target word was .083.

A different random order of presentation of study and test items was used for each second subject. There were two restrictions on a test list: A category exemplar test item was not presented in the first three test positions, and other test items from an experimental text had to appear later in the test list than the category exemplar.

Results

Means were calculated for each subject and each item in each condition, and means of these means for error proportions are shown in Table 3. The average response time (measured from presentation of the response signal) was 328.5 ms, not significantly different across experimental conditions (the range was from 319 ms to 333 ms). In all analyses below, $p < .05$.

For the primary exemplars, the hypothesis was that they would match the encoded representations of their predicting sentences sufficiently to give more errors with the predicting text than the control text and that this difference might be evident even when the primary exemplar was presented by itself as a target word with only the neutral prime. This is the pattern of data shown in Table 3: There are more errors with the predicting text in both the neutral and text word priming conditions.

For the secondary exemplars, there is no evidence that the predicting texts led to higher error rates than did the control texts. Error rates were not higher for the predicting text either with the neutral prime or the prime from the text.

Table 3
Results from Experiment 1

Condition	Neutral prime (%)	Text word prime (%)
Primary exemplars (% incorrect *yes* responses)		
Predicting text	66	63
Control text	40	51
Secondary exemplars (% incorrect *yes* responses)		
Predicting text	41	38
Control text	41	36
Filler test items		
Primary exemplar group		
Words from texts (% incorrect *no* responses)	22	18
Words not from texts (% incorrect *yes* responses)	36	37
Secondary exemplar group		
Words from texts (% incorrect *no* responses)	21	17
Words not from texts (% incorrect *yes* responses)	36	36

When the two sets of results are combined, they show an effect of predicting versus control text for the primary but not the secondary exemplars. This interaction is significant in a between-groups analysis of variance, $F(1, 46) = 10.5$, with subjects as the random variable, and $F(1, 62) = 13.6$, with items as the random variable. The main effect of primary versus secondary exemplar (different groups of subjects) was significant, $F(1, 46) = 7.7$, with subjects as the random variable, and $F(1, 62) = 14.2$, with items as the random variable, as was the main effect of predicting versus control texts, $F(1, 46) = 11.1$ and $F(1, 62) = 14.4$, respectively. No other effects were significant. However, we worried that the difference for the primary exemplars between the error rates in the predicting and control conditions with the prime from the text was too small (63% vs. 51%), so post hoc tests were performed on these two data points; the difference was significant, $F(1, 46) = 10.5$ and $F(1, 62) = 4.7$, respectively. We speculate that the prime from the text leads to a somewhat high error rate (51%) even with the control texts because of the individual words in the control text that are semantically associated with the primary exemplar (e.g., *breakfast, juice,* and *squeeze* are associated with *orange*). The standard error on the error rates was 3%.

The data from the filler test items are also shown in Table 3 and generally indicate that there were not large differences between the two groups of subjects. These data also allow a comparison between the proportions of *yes* responses for words that were explicitly stated in the texts and the primary exemplar target words that were not explicitly stated. Overall, these proportions were 80.5% and 64%. Thus, the proportion of *yes* responses for the primary exemplars is not as high as for words that were explicitly stated, but it is quite high relative to the proportion of errors (about 36%) for other test words for which the correct response was *no.*

Discussion

Interpretation of the pattern of data from Experiment 1 depends on the assumption that delayed recognition measures inference processes that occur during reading. Several considerations support this assumption. First, the inferences are not due to slow, strategic processes that occur at the time of the recognition test because response times are too fast (see also McKoon, 1988; McKoon & Ratcliff, 1986, 1988; Ratcliff & McKoon, 1981) and because the probability that any one test item will require an inference of the kind under investigation (.083) is too low to prompt strategies specific to the materials and the experimental situation. Second, the inference processes are not the backward process of compatibility checking in Forster's (1981) model. This compatibility is computed from a word to be identified back to current context, and current context in the delayed recognition procedure is not the predicting text but instead is the list of preceding test words. Third, Potts et al. (1988) proposed an alternative interpretation of the results from delayed recognition, namely that the results are due to inference processes that occur at the time of the recognition test. However, this interpretation can be ruled out. Potts et al. suggested that a prime word from a studied text activates that text, and then the target word is judged for compatibility with the activated text. By this mechanism, the response to a target word could be inhibited even if no inference about the target was made during reading. However, this reasoning predicts that *any* test word that is compatible with the text should show inhibition in the delayed text, even words that would not be inferred, and McKoon and Ratcliff (in press-a) have shown that this prediction does not hold. They demonstrated that there are test words that are compatible with a text (as judged by ratings and as measured by an on-line lexical decision task) that do not show inhibition in delayed recognition. Furthermore, compatibility ratings do not predict probability of recognition for individual items (McKoon & Ratcliff, in press-a), as Potts et al.'s hypothesis would predict. Thus, the interpretation that compatibility relations between the target and the text are calculated at the time of the recognition test can be rejected.

Having ruled out these alternative interpretations, speeded item recognition can be understood within the framework of current theories of the retrieval processes involved in recognition (Gillund & Shiffrin, 1984; Hintzman, 1986; Murdock, 1982; Ratcliff, 1978, 1988; Ratcliff & McKoon, 1988a, 1988b). These theories are supported by experimental data from a number of paradigms, including recognition, paired associate learning, free recall, frequency judgment, and categorization. The picture of recognition memory that emerges from this wide range of research is that recognition is a passive parallel process. According to the theories, the target item in its context (i.e., the target with its prime) is matched against information in memory, and the recognition response is determined by the goodness of the match. This goodness of match represents the interaction between encoded information in memory and the target in its retrieval context. Thus, if the value of goodness of match is high for an inference target (such as orange), then there must be information about that target in the mental representation of the text. Furthermore, because the match against memory represents an interaction of the target in its context and information in memory, the value of the match can be different in different contexts.

The results of Experiment 1 present a clear pattern: For the primary exemplars, the error rate is higher after the predicting text than the control text in both retrieval contexts. These data are consistent with the intuition that there is a high degree of match between these contextually defined category members and the encoded representation of the predicting text. Information that matches the primary exemplars is accessible even in impoverished retrieval conditions, as information making up highly encoded inferences should be. However, it should be noted that the data of Experiment 1 cannot measure any absolute degree of encoding; all that the experiment can actually measure is the relative degree of encoding. Information that matches primary exemplars is available in the neutral retrieval context, so this information must be encoded to a higher degree than information that matches secondary exemplars, which is not available in that context.

For the secondary exemplars, the error rate is not higher after the predicting text than the control text in either retrieval context. These data suggest that the secondary exemplars do not match the encoded inference of the predicting text any better than they match the control text. This was not a necessary result; it could have been that the secondary exemplars showed some degree of match that was lower than the match for the primary exemplars. For example, the error rate for the secondary exemplars could have been higher for the predicting texts than the control texts, but not as much higher as for the primary exemplars. Or the error rate could have been higher, but only with the prime from the test and not the neutral prime (as was the case for the inference about death for the actress; McKoon, 1988; McKoon & Ratcliff, 1986). However, neither of these patterns of results was obtained. Although it is accepting the null hypothesis to claim that there is no match at all between the secondary exemplar and the encoded inference, it must at least be the case that the match is much lower for the secondary exemplars than for the primary exemplars.

If a secondary exemplar does not match the encoded inference at all, then it can be argued that the inference does not include any significant amount of general information about the contextually defined category. Grapefruit do have many characteristics that allow them to be squeezed for juice, but this information does not appear to be available when the target word *grapefruit* is matched against the text about squeezing juice for breakfast. Instead, the results suggest that the inference about the juice being squeezed contains information specific to the primary exemplar and the text; the inference contains information about orange juice for the attorney's breakfast more than information about the general category of fruits that can be squeezed for juice. This suggestion is discussed further in the General Discussion section.

Experiment 2

If, in Experiment 1, information matching the primary exemplars received a relatively high degree of encoding during reading, then it should be possible to demonstrate the effects of this information in "on-line" comprehension. This was the purpose of Experiment 2, and an example of the materials is shown in Table 4. Sentences were presented to subjects in pairs: the first sentence labeled a context sentence and the second sentence, which followed immediately, labeled a target sentence. The variable of interest was reading time on the

Table 4
An Example of the Materials Used in Experiment 2

Predicting context
The young attorney wanted to make sure she had fresh juice for breakfast, so she squeezed the fruit herself.
Target with primary exemplar
The oranges were large, ripe, and super juicy.
Target with secondary exemplar
The grapefruit were large, ripe, and super juicy.
Neutral context
After the children left for school, the housewife went shopping and was pleased to find the fruit on sale.
Target with primary exemplar
The oranges were large, ripe, and super juicy.
Target with secondary exemplar
The grapefruit were large, ripe, and super juicy.

target sentences. The predicting context sentences were the same as the predicting sentences used in Experiment 1, and the target sentences mentioned either the primary or the secondary exemplar from Experiment 1. The neutral context sentences were written to be neutral with respect to predicting either the primary or secondary exemplar.

For the predicting sentence in Table 4, if comprehension includes generation of inferred information about *orange*, then the explicitly stated *orange* in the following target sentence should be understood as matching that information. In contrast, the neutral sentence does not point toward a particular fruit, so the *orange* in the target sentence should be understood only as an instance of a more general category of fruits (those for which you can go shopping). These hypotheses can be examined by measuring the reading times of the target sentences. The idea is that if the primary exemplar matches information inferred during reading of the predicting sentence but not the neutral sentence, then reading time for the target sentence with the primary exemplar should be faster following the predicting sentence than the neutral sentence.

For target sentences with the secondary exemplar, the results of Experiment 1 lead to a different prediction. The secondary exemplar does not have a high degree of match with the inference generated by the predicting sentence. So, for the secondary exemplar, there may be little advantage to the predicting context over the neutral context; that is, there may be little difference in reading time for the target sentence in the two contexts.

Using the reading times of target sentences to support theoretical predictions requires making several assumptions about how those target sentences are processed. For example, it must at least be assumed that the context and target sentences are read as part of the same discourse. Discussion of these sorts of assumptions can be focused by empirical results, and so their presentation is postponed until the *Discussion* section below.

Method

Subjects and design. There were two separate groups of subjects. For one group of 18, the target test sentences always contained the primary exemplar; for the other group of 20, the targets always contained the secondary exemplar. For each group, there were two conditions, so that the context sentence that preceded the target was either the predicting sentence or the neutral sentence. The two conditions were combined in a Latin-square design with groups of subjects (9 or 10 per group) and sets of materials (16 per set). The subjects participated for credit in an introductory psychology class.

Materials. For each of the 32 predicting texts used in Experiment 1, a neutral context sentence and two target sentences were written. One target sentence mentioned the primary exemplar from Experiment 1, and the other mentioned the secondary exemplar. The average length of the target sentences was 8.9 words, and the average length of the neutral context sentences was 23.4 words.

The neutral context sentences were not the same as the control sentences of Experiment 1. Those control sentences were written to bias the interpretation of the category away from the primary and secondary exemplars so that inferences due to the meaning of a whole text could be most clearly separated from inferences due to the meanings of the individual words associated with the test word. In Experiment 2, the neutral context sentences could not bias the interpretation away from the primary and secondary exemplars; otherwise, the resulting context–target pairs would be somewhat odd (and reading times would be slowed, as Roth & Shoben, 1983, showed). For example, the control sentence in Table 1 mentions a fruit that could be hard and green, a description that would not usually be applied to an orange. Instead, the neutral sentences were written to provide as nearly as possible a genuinely neutral context (e.g., fruits you can shop for; see Table 4). In this way, advantages given to target reading times by the predicting context would be most compelling: They can be more easily interpreted as facilitation given by the predicting sentence and not as inhibition due to a control sentence that biases against the target.

There were also 32 filler texts, each with two sentences (the first averaged 20.7 words, the second 10.9 words). For each of these a true–false test question was written. These questions were used so that subjects would be more likely to attend carefully when reading the texts.

Procedure. Each subject participated in one 50-min session. As in Experiment 1, all materials were presented on a CRT screen, and responses were collected on the CRT's keyboard.

There were 8 texts for practice, and then 64 texts for the experiment proper. Each text began with an instruction on the CRT screen to press the space bar on the keyboard. After a 100-ms blank interval, the first sentence of the text was displayed. The sentence remained on the screen until the space bar was pressed again; then the sentence disappeared, there was a 50-ms pause, and the second sentence of the text was displayed in the same location. The second sentence again remained on the screen until the space bar was pressed. Then after a 200-ms pause, the message to press the space bar for the next text was displayed. Subjects were instructed to read each sentence carefully, pressing the space bar when they felt they had understood the sentence. The texts were presented in sets of eight, four from the experimental set and four fillers. After each set of eight, there were four true–false test questions. These were signaled by a message on the CRT screen. Subjects were instructed to respond as accurately as possible to these questions, pressing the ?/ key for *true* and the Z key for *false*. If a response was incorrect, the word ERROR was displayed for 2,000 ms. Otherwise, each true–false question followed 500 ms after the last. Order of presentation of texts was random and different for every second subject.

Results

Average reading times for the target sentences were calculated for each subject and each item in each condition. For the primary exemplar target sentences, reading times were faster with the predicting context than the neutral context (2,256 ms vs. 2,482 ms). This difference was significant with subjects as a random variable, $F(1, 17) = 7.4$, and with sentences as the random variable, $F(1, 31) = 4.7$. For the secondary exemplar target sentences, there was no difference in reading times (2,482 ms vs. 2,403 ms) for the predicting and neutral conditions, respectively (Fs both < 1.0). The MS_e on these reading times was 59 ms for both primary and secondary exemplars.

Average reading times were also obtained for the context sentences. For the group of subjects who were tested with the primary exemplar targets, reading times for the predicting sentence averaged 5,749 ms, and reading times for the neutral sentence averaged 5,629 ms. The comparable times for the group with secondary targets were 5,545 ms and 5,314 ms.

For the filler texts, reading times were recorded for both the first and second sentences. For the primary exemplar group of subjects, the average times were 5,472 ms and 2,622 ms, first and second sentences, respectively, and for the secondary exemplar group, 5,166 ms and 2,519 ms. The error rates on the true–false test questions were 5% on true sentences and 8% on false sentences for the primary exemplar group and 7% and 6% for the secondary exemplar group.

From the data for the context sentences and the filler sentences, the subjects who were tested with the secondary targets appear to have read more quickly than the subjects tested with the primary targets. It might be that this fast reading speed led to the absence of a significant effect of predicting versus control on target reading times. So the data for the eight slowest subjects in the secondary group were analyzed separately. The mean for the targets in the predicting condition was 2,994 ms, compared with 2,887 ms in the neutral condition. Thus, even for the slowest subjects, there was no effect of predicting versus neutral context for the secondary exemplar targets.

Discussion

The results of Experiment 2 were exactly as expected from the results of Experiment 1: A context sentence that defined a category facilitated reading time for a target sentence that mentioned the primary exemplar of the category but did not facilitate reading time for a target that mentioned a secondary exemplar. The obvious interpretation of these results is that information matching the primary exemplar was inferred during reading of the predicting context sentence, and the match between inferred information and the explicitly stated exemplar in the target sentence facilitated reading time for the target sentence. The secondary exemplar did not match the inferred information, and so reading time was not facilitated.

This interpretation requires assumptions that need to be made clear. One is that the explicitly stated concept (e.g., orange) in the target sentence is understood in terms of the previously mentioned fruit in the context sentence for both the predicting and the neutral context sentences. The second is that processing the connection between an explicitly stated concept (e.g., orange) and information about that concept in an inference is faster than connecting the explicitly stated concept to a more general category (e.g., fruits you can shop for).

These assumptions are most easily explained by supposing that they are not true. First, suppose that for the neutral context, the explicitly stated concept in the follow-up target sentence was not understood as an instance of the general category mentioned in the neutral sentence. Instead, it was treated as a new concept. Then slow reading times for the target sentence in the neutral context could be due to slow processing for this new concept (Kintsch, Kozminsky, Streby, McKoon, & Keenan, 1975). All that would be required to give faster reading times for the target sentence in the predicting context would be for the predicting sentence to give some general inference about the category (e.g., fruits to shop for), and not any specific inference about the primary exemplar (e.g., oranges). To rule out this interpretation, it must be assumed that the oranges mentioned in the target sentence are understood to refer to the category defined in the context sentence for both the predicting and the neutral contexts.

The second assumption is that it is faster to match the explicitly stated oranges in the target sentence to a specific inference (from the predicting context) than to a more general inference (from the neutral context). If this were not true, if the two processes required equal amounts of time, then differences in target sentence reading times would have to be due to some other differences between the conditions. There are many possible differences between pairs of sentences, ranging from vocabulary to style to complexity, and they cannot all be controlled, even though obvious possible confounds were eliminated when the materials were written.

These two assumptions are not trivial, and they cannot be defended from the data of Experiment 2 alone. But, on the other hand, it is plausible that an explicitly stated exemplar of a category would be understood as referring to an immediately preceding description of that category, and it is plausible that connecting an explicitly stated concept to specific inferred information would be faster than connecting it to more general information. Besides plausibility, the assumptions can also be defended by drawing analogies to previous work. A number of studies have used more standard categories than the contextually defined categories used here. For example, Sanford and Garrod (1977) examined reading times for sentences that mentioned an exemplar of a standard category that had been previously stated (e.g., *oak* when the category *tree* had been stated). In all of this work (Corbett, 1984; Dell, McKoon, & Ratcliff, 1983; McKoon & Ratcliff, 1980; Sanford & Garrod, 1977), variables that govern the relation between the exemplar and the category (such as typicality) affect reading time for the target sentence. Thus, to provide a consistent interpretation across these studies, it is assumed that in Experiment 2 the exemplars were understood in terms of the previously stated categories and that more appropriate or typical exemplars speeded comprehension.

General Discussion

In summarizing the experiments of this article, the results of the delayed recognition experiment (Experiment 1) and the results of the reading time experiment (Experiment 2) are considered in combination along with the contrast between the primary and secondary exemplars. Then general conclusions are discussed about the contextual definition of categories and about inference processes.

To begin, the results suggest that information matching the most typical exemplars of contextually defined categories is encoded to a relatively high degree. This high degree of encoding makes it difficult, in delayed recognition, for subjects to correctly determine that the exemplar was not explicitly presented in its text. Subjects make a large proportion of errors, deciding that the exemplar was in the text when in fact it was not, and they do so even when the exemplar is presented by itself with only the neutral prime. It is this result, that the exemplar matches the mental representation of the text so well by itself, that indicates a high degree of encoding.

If information matching the primary exemplar is encoded to such a high degree when the contextual description of the category is read, then a subsequent, explicit mention of that exemplar in a follow-up sentence should be easy to comprehend. This prediction was confirmed in Experiment 2: The contextual description of the category was presented in a context sentence immediately preceding a target sentence that mentioned the primary exemplar. Reading time for the target was speeded relative to when it followed a neutral context sentence. Thus, the results from the delayed recognition experiment (Experiment 1) and the reading time experiment (Experiment 2) provide converging evidence that information matching the most typical exemplars of contextually defined categories is encoded to a high degree, supporting Roth and Shoben's (1983) earlier conjectures.

In contrast, the results of the experiments indicate that other, secondary exemplars of contextually defined categories do not match the information encoded about the category. In delayed recognition, subjects had no significant difficulty in correctly rejecting the secondary exemplars, that is, in deciding that they had not been presented in a text. This was true even when the secondary exemplar was combined with a prime from the text, indicating that the exemplar had no match at all (or very little match) with the mental representation of the text.

If a contextual description of a category does not lead to the encoding of information that matches secondary exemplars, then there should be no facilitation of comprehension for a subsequent, explicit mention of that exemplar. This prediction was confirmed in Experiment 2 when target sentences with secondary exemplars were read no faster following context sentences that described their category than following neutral context sentences. As with the predicting exemplars, the results of Experiments 1 and 2 converge.

The contrast between the results for the primary and secondary exemplars gives considerable leverage with which to look at inferences about contextually defined categories. Conclusions can be drawn both about the content of these inferences and about the processes that construct them. First, hypotheses about content will be considered. The data provide two restrictions: The encoded features must match a primary exemplar very well, and they must not match a secondary exemplar at all (or very little).

If the encoded inference does not match a secondary exemplar, then the inference cannot be made up of the defining and characteristic features of the category. Although this idea would be consistent with adapting traditional network models of semantic memory to deal with contextually defined categories (Anderson, 1976; Collins & Loftus, 1975; Roth & Shoben, 1983), it could not be used to account for the data because it would predict that secondary exemplars would match the inference to some degree. For example, there would be many features of grapefruit that matched features of the category *fruits you squeeze for juice for breakfast.* But both the delayed recognition data and the reading time data indicate that there is little, if any, match.

The problem posed by the data is to explain why the encoded features of the inference match the primary exemplar so well while not matching the secondary exemplar at all. There are several solutions to this problem, all of which involves some kind of instantiation of the primary exemplar in the inference (Roth & Shoben, 1983). One possibility is that the primary exemplar is encoded almost as though it had actually been read, except that perceptual information (e.g., about letter fonts or position in the sentence) would be lacking. The encoded meaning of the inference would be the same as if the word were read, and this meaning would not match the secondary exemplar. Thus, the encoded meaning would relate the primary concept (e.g., orange) and features specific to that concept to the other information in the text, about the attorney, breakfast, and so forth.

Another solution to the problem is to assume that the inference is made up of the features that define and characterize the contextually defined category but that these features are not encoded as general features. Instead, they are tied to the primary concept in such a way that the secondary concept cannot match them. For example, the general feature *contains Vitamin C* would not be encoded; instead, the information that oranges contain vitamin C would be encoded. This solution is not appealing because it requires one kind of mismatching information (*grapefruit* would mismatch *oranges*) to override another kind of matching information (the *contains Vitamin C* feature of *grapefruit* would match the encoded information *contains Vitamin C*).

A third way to explain why the encoded inference matches the primary exemplar so well is to assume that the inference represents a sample of instances of having juice squeezed from a fruit for breakfast (Hintzman, 1986; Jacoby & Brooks, 1984; Medin & Shaffer, 1978). These instances would mostly contain oranges because they are, by far, the most frequent fruits used for breakfast juice in most people's experience. Thus, the consensus of the instances would be features that exactly describe oranges and not grapefruit. The inference would reflect these features and so orange would match well. But grapefruit would not match because many of the features would be specific to oranges (e.g., a particular size and color).

These three different ways of describing instantiation of the primary exemplar are speculation; current data cannot distinguish among them. However, the data of Experiments 1 and 2 do allow rejection of the hypothesis that general information about the contextually defined category is encoded and point toward some kind of instantiation of the primary exemplar.

The instantiation hypothesis leads straightforwardly to the explanation of the delayed recognition data from Experiment 1: The primary exemplar figures prominently in the inference, and so, when it is presented as a test word, the degree of match is high. The secondary exemplar is represented in the inference to some very low degree, if at all, and so its degree of match is low.

Given this picture of the content of the inference, various hypotheses about processing of the context and target sentences of Experiment 2 can be considered. First, it might have been the case that no inference about the contextually defined category was formed during reading of the context sentence. Inferences were formed only during reading of the target sentence when the explicit mention of the primary exemplar in the target sentence and the contextual description in the context sentence could be combined. However, this hypothesis can be rejected because the delayed recognition data of Experiment 1 indicate that the context sentence alone is sufficient to give the inference.

Another hypothesis might have been based on the notion that the content of the inference represented a general description of the features of the category. Then when the primary exemplar was explicitly mentioned in the target sentence, it could be understood as having those features, and the secondary exemplar could also be understood as having the features, but with extra processing. But this hypothesis would predict some degree of match in delayed recognition between the secondary exemplar and the mental representation of the context sentence, and the data failed to show such a match.

The hypothesis that is most consistent with the data from both the delayed recognition and reading time experiments and from both the primary and secondary exemplars is that the comprehension processes are different for the target sentence with the primary exemplar than for the target sentence with the secondary exemplar. If the content of the inference about the category is constructed during reading of the context sentence and if that content instantiates the primary exemplar in some way, then the explicit mention of the primary exemplar can be comprehended simply by connecting it to the information in the inference. But the secondary exemplar cannot be understood in this way because there is little (or no) information in the inference with which it can be connected. Instead, comprehension of the secondary exemplar in the context of the contextually defined category must require extra processing. What this extra processing might be is speculation at this point: It might take the form of reconstructing the inference so that it does represent a list of defining and characteristic features to which the secondary exemplar can be fit. Or it might take the form of checking that the secondary exemplar is an acceptable exemplar of the category (cf. Becker, 1980; Roth & Shoben, 1983) and then substituting it for the primary exemplar in the features or propositions making up the inference.

To summarize, when a category is contextually defined, then information about a most typical exemplar of the category is encoded to a high degree of inference processes. The encoded information is then available to be connected to later mentions of the exemplar, or it can be changed at a later time to allow other exemplars to be substituted. It should be stressed, however, that these processes can be claimed to apply only for contextually defined categories that have a most typical exemplar, as was the case for the categories used in the experiments in this article. Other categories may have so little definition that there is no most typical exemplar (e.g., *Dorothy thought about the animal,* from Roth & Shoben, 1983). Thus, there would be no high degree of encoding for a specific exemplar.

This article began with discussion of a proposal that inferences could be encoded to varying degrees, ranging from not encoded at all to actually instantiated. The purpose of the experiments was to investigate one variable that might determine the degree of encoding for a particular kind of inference. This variable was the amount of information (or number of associations) in long-term memory that connected the explicitly stated information in a text to the information to be inferred from long-term memory. This variable was implemented by comparing the most typical exemplars of contextually defined categories with less typical exemplars. The most typical exemplars should have many long-term memory associations to connect them to the contextually defined category, whereas the less typical exemplars should have relatively few. The consequence of this variation in amount of long-term memory information was that information about the typical exemplars was encoded to a high degree, and information consistent with other exemplars was not. Thus, the experiments supported the proposal that inferences can be encoded to varying degrees.

Experiments 1 and 2 also support a general framework of minimal inference processing (McKoon & Ratcliff, in press-b). This view suggests that if an inference is not required to give coherence to the information in a text, then the inference will not be constructed. The conclusions of this article modify this view: If a specific inference is provided by easily available general knowledge from long-term memory, then it will be constructed even if it is not required for coherence. But this modification applies only for the most easily available information; other inferences, such as those about predictable events (McKoon & Ratcliff, 1986; Singer & Ferreira, 1983), about default values in schema representations, (Alba & Hasher, 1983; Seifert, McKoon, Abelson, & Ratcliff, 1986), and about the instruments of verbs (Corbett & Dosher, 1978; McKoon & Ratcliff, 1981), are usually encoded partially or not at all.

References

Alba, J. W., & Hasher, L. (1983) Is memory schematic? *Psychological Bulletin, 93,* 203–231.

Anderson, J. R. (1976). *Language, memory, and thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, R. C., & Ortony, A. (1975). On putting apples into bottles—A problem of polysemy. *Cognitive Psychology, 7,* 167–180.

Becker, C. A. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory & Cognition, 8,* 493–512.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82,* 407–428.

Corbett, A. T. (1984). Pronominal adjectives and the disambiguation of anaphoric nouns. *Journal of Verbal Learning and Verbal Behavior, 17,* 683–695.

Corbett, A. T., & Dosher, B. A. (1978). Instrument inferences in sentence encoding. *Journal of Verbal Learning and Verbal Behavior, 17,* 479–491.

Dell, G. S., McKoon, G., & Ratcliff, R. (1983). The activation of antecedent information during the processing of anaphoric reference in reading. *Journal of Verbal Learning and Verbal Behavior, 22,* 121–132.

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *Quarterly Journal of Experimental Psychology, 33,* 465–495.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 19,* 1–65.

Hintzman, D. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: memory, perception and concept learning. In G. H. Bower (Ed.), *The psychology of learning and motivation theory: Advances in research and theory.* (Vol. 18, pp. 1–47). New York: Academic Press.

Kintsch, W., Kozminsky, E., Streby, W. J., McKoon, G., & Keenan, J. M. (1975). Comprehension and recall of text as a function of content variables. *Journal of Verbal Learning and Verbal Behavior, 14,* 196–214.

McKoon, G. (1988). *Word identification and elaborative inference: The relation between theory and empirical measurement.* Manuscript submitted for publication.

McKoon, G., & Ratcliff, R. (1981). The comprehension processes and memory structures involved in instrumental inference. *Journal of Verbal Learning and Verbal Behavior, 20,* 671–682.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 82–91.

McKoon, G., & Ratcliff, R. (1988). Contextually relevant aspects of meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 331–343.

McKoon, G., & Ratcliff, R. (1989). Semantic association and elaborative inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 326–338.

McKoon, G., & Ratcliff, R. (in press-a). *Assessing the occurrence of elaborative inference with recognition: Compatibility checking vs. compound cue theory. Journal of Memory and Language.*

McKoon, G., & Ratcliff, R. (in press-b). Dimensions of inference. *The psychology of learning and motivation.*

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626.

Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inference: Lexical decision versus naming. *Journal of Memory and Language, 27,* 399–415.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling the accumulation of partial information. *Psychological Review, 95,* 238–255.

Ratcliff, R., & McKoon, G. (1981). Automatic and strategic priming in recognition. *Journal of Verbal Learning and Verbal Behavior, 20,* 204–215.

Ratcliff, R., & McKoon, G. (1988a). Memory models, text processing, and cue-dependent retrieval. In F. I. M. Craik & H. L. Roediger (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 73–92). Hillsdale, NJ: Erlbaum.

Ratcliff, R., & McKoon, G. (1988b). A retrieval theory of priming in memory. *Psychological Review, 95,* 385–408.

Roth, E. M., & Shoben, E. J. (1983). The effect of context on the structure of categories. *Cognitive Psychology, 15,* 346–378.

Sanford, A. J., & Garrod, S. C. (1977). Memory and attention in text comprehension: The problem of reference. In R. S. Nickerson (Ed.), *Attention and performance VIII.* Hillsdale, NJ: Erlbaum.

Seifert, C. M., McKoon, G., Abelson, R. P., & Ratcliff, R. (1986). Memory connections between thematically similar episodes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 220–231.

Singer, M., & Ferreira, F. (1983). Inferring consequences in story comprehension. *Journal of Verbal Learning and Verbal Behavior, 22,* 437–448.

Smith, E. E., Shoben, E. J., & Rips, J. L. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review, 81,* 214–241.

Appendix

Materials Used in Experiment 1

For each item, the text marked *a* was the predicting text, and the text marked *b* was the control text. The first word following the texts was used as the most typical exemplar, and the second word was used as the less typical exemplar.

1a. The young attorney wanted to make sure she had fresh juice for breakfast, so she bought and squeezed the fruit herself.

1b. When the waiter brought breakfast, the attorney squeezed the fruit to make sure it was juicy and ripe, not hard and green. *orange, grapefruit*

2a. The old man loved his granddaughter, and she liked to help him with his animals; she volunteered to do the milking whenever she visited the farm.

2b. The old man loved his granddaughter, and she liked to help him with his animals when she visited the farm; she also liked the milk and cookies her grandmother provided. *cow, goat*

3a. In the final scene of the play, the actress stabs her lover and, as the police are about to arrive, desperately tries to hide the weapon.

3b. In the final scene, the actress threatens to shoot her lover with his own weapon when suddenly she feels the barrel of the police revolver stab her from behind. *knife, scissors*

4a. Waving her net, the little girl giggled as she followed the brightly colored wings around the garden, trying to catch the fluttering insect in the net.

4b. The little girl giggled as she chased her brother behind the fluttering volleyball net, but she started to cry when an insect stung her. *butterfly, moth*

5a. Telling the driver to keep the change, the businessman grabbed his briefcase, got out, and slammed the door.

5b. The doors closed as the businessman climbed aboard, and the driver told him he must have exact change. *taxi, limousine*

6a. Susie pulled out the bench, sat down, and called to everyone in the family to gather round: "It's Christmas—I'll play and we'll all sing carols."

6b. At the Christmas service, sitting on a bench near the altar, Susie strummed carols and encouraged the congregation to sing along. *piano, organ*

7a. The retired grandmother amazed everyone; last winter when she went up to a mountain resort was the first time she tried the sport.

7b. There were indoor courts at a nearby mountain resort, so the grandmother played all winter, amazing everyone with her energy for sports. *skiing, skating*

8a. While the movers took a break, Betty went to her room for a quick nap, thankful that at least one piece of furniture had not been loaded yet.

8b. Betty was furious to find one of the movers taking a nap in the cab of the van, especially since she had just discovered a scratch on her most prized piece of furniture. *bed, couch*

9a. This Thanksgiving, mother had really outdone herself; the bird highlighting the meal was plump and juicy, roasted to perfection.

9b. During Thanksgiving vacation, mother took the children outside to get some exercise and feed the birds some stale stuffing crumbs. *turkey, pheasant*

10a. Cup after cup, with no milk or sugar; you'd think the secretary had to keep drinking to stay awake until the end of the morning.

10b. It was the secretary's first meeting with the boss and an important occasion, so she kept drinking to have something to do with her hands. *coffee, tea*

11a. The connoisseur kept the bottles in the cellar for years, knowing that he would fully appreciate drinking them when the occasion was right.

11b. The toddler had managed to get into the kitchen with no one around; the worst of the damage was a bottle lying broken on the floor. *wine, champagne*

12a. Although the salesman tried to convince Alan that the right kitchen appliance could change his life, Alan just wanted something to keep his food cool.

12b. Alan complained to the appliance salesman that the electronic marvel he had sold him didn't work on all heat settings and left parts of his food cool. *refrigerator, freezer*

13a. Margaret was usually not interested in jewels, but the one that sparkled from her engagement ring was so brilliant that she couldn't believe her fiance had bought it.

13b. Margaret didn't believe that her fiance had actually gone diving in the Pacific for the jewel in her engagement ring, but she liked the glow and faint sparkle so much, she didn't care. *diamond, ruby*

14a. The boys were so excited that they were finally allowed to have a pet that they used their allowance to buy a collar, and bowls for food and water.

14b. The boys were excited to get their allowance money and used it to buy a collar for their pet hamster, and also got some new bowls for food and water. *dog, cat*

15a. The dentist was running behind today, so the patient sat down in the waiting room and picked up something to read.

15b. While the patient waited in the outer office, the dentist hurriedly read the X-ray; he was running far behind. *magazine, newspaper*

16a. The tour of the city included many historic places and buildings; an especially emotional event was attendance at religious services on Sunday, where the atmosphere was hushed and the people were solemn.

16b. In his course on Western philosophy and religion, the freshman excelled at multiple-choice questions about dates, events, places, and buildings, but could write very little in the way of creative commentary. *church, temple*

17a. Pointing at the basket of fruit, the baby demanded "the yellow one," so the babysitter peeled the fruit and mashed it up.

17b. The baby grabbed the box of dried fruit from the babysitter and tried to mash it up; he just couldn't wait for dessert. *banana, apple*

18a. The professor wanted to sample the culture, so when he arrived in Russia, he was anxious to try the local liquor.

18b. Sipping an after-dinner liquor, the professor who had just returned from Russia told everyone tales of the local culture. *vodka, beer*

19a. Although she usually liked animals, the housewife was annoyed with the furry little beasts who got into her garden and fed themselves on her lettuce.

19b. The housewife lived near the stables, so when she picked lettuce from her garden, she liked to go over and feed the animals that were kept there. *rabbits, gophers*

20a. The concert master was nervous because the Mozart piece was difficult; she tuned and retuned, checked for extra strings in case she broke one, and adjusted her music.

20b. Ellen had been complaining that she couldn't play well because of worn strings and hammers, so she was glad when the man from the concert hall showed up to do some repairs. *violin, cello*

21a. After concentrating on his opening move for several minutes, Boris finally moved a piece across the board to begin the game.

21b. Moving the soldiers and battle weapons around the board and planning the opening attack on the enemy, Boris reminded himself that it was just a game. *chess, checkers*

22a. The lifeguard spotted the dangerous tail fin, but before he could act, several swimmers started running from the water, creating panic on the beach.

22b. When the swimmers told the lifeguard about a fish with the body of a woman, fins, and a tail, he warned them not to panic the creature. *shark, barracuda*

23a. At church on Easter Sunday, a bouquet of Easter flowers contributed by the church ladies group was proudly displayed at the altar.

23b. It was Easter Sunday, and although she knew it was just a weed, the girl picked the flower to wear to church. *lily, daffodil*

24a. The student hated dirt and kept his kitchen spotlessly clean, but there was nothing he could do about the insect problem; the invasion of his kitchen every night had gotten out of control.

24b. The student stomped out of the kitchen, trying to get away from the insects that kept buzzing around his food and annoying him; he would have to buy some window screens to solve the problem. *cockroach, ant*

25a. When the gang got together for their game last week, the foreman seemed to hold all the winning cards and was well on his way to paying off all his debts.

25b. The foreman needed to pay off his debts, so before betting on the game, he checked the statistics cards and realized that Johnson was no longer a winning pitcher. *poker, rummy*

26a. Irv asked his wife to make his favorite vegetable, but she claimed that they took too long to bake and contained far too much starch for his own good.

26b. Irv complained to his wife that her baked lasagna had far too much starch in it, and he chomped on raw vegetables to keep to his diet. *potatoes, yams*

27a. David realized that in his profession he was responsible for people's health and even their lives, but initially he had only been interested in making money.

27b. David knew that his profession was less than honorable since it consisted of threatening people's health and taking their lives for money, but he was very good at it. *doctor, lawyer*

28a. The detective hoped he could find the criminal; the blood stains were everywhere and indicated such a vicious mind that he worried about the possibility of future victims.

28b. The blood stains showed the police detective where the guard dogs had been killed, enabling the criminal to escape with the loot. *murder, kidnapping*

29a. At the camp, the city boys loved the riding, even though they had to learn to care for the animals every day.

29b. The city boys liked to ride the subways to keep count of the number of animals they saw scurrying through the dark. *horses, donkeys*

30a. The kindergartner stood crying in alarm; his pet "Leo" was swimming frantically around his bowl, terrified of the watching cat.

30b. The kindergartner stood crying; his pet cat "Leo" had followed his bowl when it fell into the swimming pool; Leo was frantically thrashing around. *fish, turtle*

31a. When the alarm went off, it woke the householder; he grabbed his children and ran downstairs and out the front door to escape the emergency.

31b. The householder was a doctor, and whenever his beeper alarm went off, it meant an emergency at the hospital. *fire, tornado*

32a. Deep red in color and extremely fragrant, it was no wonder that the flower, named American Beauty, won the competition.

32b. With red hair, fragrant perfume, and flowers, the girl easily won the beauty competition and got a scholarship for college. *rose, carnation*

Received July 6, 1988
Revision received February 1, 1989
Accepted February 2, 1989 ■