

Meaning, amount of processing, and memory for words

P. N. JOHNSON-LAIRD

University of Sussex, Brighton, England BN1 9QG

G. GIBBS

Open University, Walton Hall, Milton Keynes, England

and

J. de MOWBRAY

University of Sussex, Brighton, England BN1 9QG

Two experiments were carried out in order to try to resolve the controversy about whether the semantic processing of a word necessarily involves all the elements of its meaning. In the first experiment, subjects categorized a list of auditorially presented words according to whether or not they were natural consumable solids (e.g., "apple"). They were then given an unexpected test of their ability to recall the whole list, which contained equal numbers of words with none, one, two, or all three of the target components. The results confirmed the prediction that the greater the number of components a word has in common with a target category, the better it will be recalled. The second experiment used a visual presentation of a list of words, which subjects scanned for members of a particular target category. There were four different target categories assigned to independent groups of subjects. The results again confirmed the prediction. On the assumption that memorability in such tasks is determined by amount of processing, we conclude that some elements of the meaning of a word can be processed without having to process all of them.

Semantic theories generally assume that the meanings of words contain separate components of information (see, e.g., Collins & Quillian, 1972; Miller & Johnson-Laird, 1976; Rips, Shoben, & Smith, 1973; Schank, 1972). But, what happens when you encounter a word in a sentence? Do you retrieve all of its semantic components or only some of them? Various theorists, such as Gibson (1971), have argued that when a word is processed for meaning all of its semantic components are necessarily recovered, and a number of experimental findings are compatible with this hypothesis. Hyde and Jenkins (1969) found that subjects memorizing words with the additional task of rating them as pleasant or unpleasant recalled the words just as well and in just as organized a fashion as control subjects who had no additional task to perform. Likewise, Hyde (1973) found that rating words as both pleasant or unpleasant and active or passive yielded no better recall of them than rating them on only one of these dimensions. Other results, however, support the view that usually only some semantic components of a word are processed. Klein and Saltz (1976) found that words rated on two dimensions tended to be better recalled

than words rated on only one dimension. Frase and Kammann (1974) found that recall was better when subjects categorized words in terms of specific classes (e.g., big, small) than when they categorized them in terms of general classes (e.g., animal, plant).

Unfortunately, none of these studies provides a definitive answer to our question. Tasks such as rating items as pleasant or unpleasant tap genuine semantic components for only a small proportion of words: "Unpleasant" is not, for example, part of the meaning of "lion" or else it would be as anomalous to assert that a particular lion was pleasant as to assert that a particular person was an unmarried wife. Moreover, as Klein and Saltz (1976) themselves point out, two scales yield a larger number of separate categories of words than one scale, and a subject might recollect these categories and use them as a cue to recall (see Mandler, 1967; Ornstein, Trabasso, & Johnson-Laird, 1974). Similarly, the distinction that Frase and Kammann (1974) draw on plausible but intuitive grounds between specific and general classes of words has sometimes been found to be misleading (cf. Rips et al., 1973).

In order to resolve the controversy, it is plainly necessary (1) to employ genuine semantic components of words, (2) to manipulate them in an explicitly specifiable way, and (3) to insure that the number of

categories of words is held constant. We accordingly used a task in which subjects had to detect members of a target category in a list of words of various sorts. Target categories were defined in terms of a number of semantic components, and the lists contained words with all, some, or none of the target components. For example, if the target category is consumable liquids, then "beer" is a member of it, "petrol" and "cake" each have only one of the target components, and "coal" has neither of them. We assumed that the amount of processing of a word in such a task determines its memorability, and that this variable is best indexed by the number of decisions about a word that yield pertinent information. If the semantic components of a word can be separately processed, a word with none of the target components will need only a small amount of processing in order to classify it: The decision can be made on the basis of any of its components. A word with several of the target components will generally require a greater amount of processing: The decision cannot be made on the basis of just any component, and it may be necessary to process several before encountering one that fails to match the target. In short, if amount of processing determines memorability, then it follows that the more components a word has in common with the target category, and hence the more components that may have to be checked, the better it will be remembered.

EXPERIMENT 1

Method

In this preliminary experiment, a group of subjects listened to a list of words and classified each of them as denoting either a positive or negative instance of the category of substances that are consumable, solid, and natural. The list contained four sorts of words: target words with all three components (e.g., "apple"), words with two of the components (e.g., "coal"), words with only one of the components (e.g., "sweat"), and words with none of the components (e.g., "paraffin"). After the subjects had completed the classification task, they were given an unexpected memory test in which they had to recall as many words as possible from the list as a whole.

Materials and Procedure. There were nine words in each of the four categories (zero, one, two, or all three target components), and they were balanced for length and for frequency of use (Kucera & Francis, 1967); in order of presentation, they were: paraffin, bread, coal, bronze, apple, milk, carrot, perfume, paint, mutton, soup, and so on. They were presented over a loudspeaker in a random order at a rate of one every 4 sec; pilot studies had shown that this procedure allowed the majority of subjects just enough time to classify them. The list began with three practice words, in order to allow subjects to become familiar with the task; it ended with three other such items to eliminate any special advantage to the last few test words. Each of the spoken words was preceded by its number in the list, and the subjects wrote down their classification responses on a correspondingly numbered sheet. On completing their classifications, they did mental arithmetic problems for 90 sec in order to increase the amount of forgetting, and finally, they were given an unexpected recall test in which they wrote down in any order as many of the words as they could remember.

Subjects. A group of 31 subjects was tested, but the results of 8 subjects were excluded from the analysis because they either made more than 10% errors in classification or anticipated the recall test.

Results and Discussion

The overall recall of the four sorts of words was as follows: 57.0% of the words in the target category, 44.9% of the words with two components of the target, 34.3% of the words with one component of the target, and 32.4% of the words with no component of the target. This trend was very reliable (Page's $L = 640$, $p < .0001$). Because some special advantage in memorability may be conferred upon items in the target category, we examined the trend over the three nontarget categories: It was also reliable (Page's $L = 297$, $p < .002$).

Although the results were in accord with the prediction, there are weaknesses in the experiment. It used only a single semantic category as target, and only a single order of presentation of the words. We therefore decided to carry out a second experiment designed to eliminate these defects and to extend the results to a task that involved reading.

EXPERIMENT 2

Method

There were four separate groups of subjects, each assigned to a different target category: consumable liquids, consumable solids, nonconsumable liquids, and nonconsumable solids. The same list of words was used for each group; it contained, in addition, four other categories of nontarget words: utensils used for consumable liquids (e.g., "jug"), utensils used for consumable solids (e.g., "plate"), utensils used for nonconsumable liquids (e.g., "vase"), utensils used for nonconsumable solids (e.g., "hammer").

We predicted that words would tend to be recalled as a function of the number of components that they had in common with the target category; hence, the recall of substances would be better than the recall of utensils, but there would be the same partial trend within both sorts of words. Where the target words were consumable solids, for example, they should be better recalled than either consumable liquids or nonconsumable solids, which in turn should be better recalled than nonconsumable liquids. Likewise, utensils for consumable solids should be better recalled than those for consumable liquids or nonconsumable solids, which in turn should be better recalled than those for nonconsumable liquids.

Materials and Procedure. The list consisted of eight words in each of the eight categories (e.g., sherry perfume, apple, concrete, kettle, toaster, bucket, hammer, etc.). It was presented in the form of a mimeographed booklet, with 8 words/page typed in uppercase, 1 to a line. The words were randomly assigned to the pages, and the order of the pages was randomized in each booklet.

The subjects were assigned in rotation to one of the four groups and tested individually. They were told to search rapidly through the booklets and to tick each word in the target category. When they had completed this task, they were given an unexpected recall test in which they had to write down as many of the words on the list as they could remember.

Subjects. Twenty subjects, students and nonstudents, participated in the experiment. None of them had a background

Table 1
The Percentages of Words Correctly Recalled

	Semantic Components of the Target Category Possessed by the Words			
	Both	One	Neither	Overall
Substance Words	50.0	21.5	10.6	27.4
Utensil Words	16.2	10.6	8.1	11.7

in psychology and none had participated in an experiment of this sort before.

Results

Since the four target categories had no reliable effect on the overall numbers of words recalled [$F(3,16) = 1.75, p > .1$], their results were pooled for subsequent analysis. Table 1 presents the percentages of substance and utensil words recalled as a function of the number of components of the target category that they possessed. The substance words were recalled better than the utensil words by all 20 subjects ($p = .52^0$). The predicted trend was reliably confirmed both for the substance words (Page's $L = 268, p < .0001$) and for the utensil words, even though there was probably a "floor" effect here (Page's $L = 254.5, p < .011$). The correlation between the times taken to search through the booklets and the numbers of items correctly recalled was small and insignificant ($r = .11, p > .2$).

GENERAL DISCUSSION

There can be little doubt that after listening or looking for words of a particular category, a subject is likely to recollect having encountered those that are similar rather than dissimilar in meaning to a required target. This finding suggests that a listener or reader is certainly capable of processing only some of the semantic components of a word: Once a given word has been categorized, any processing of further semantic components is unnecessary. The results cannot be explained in terms of the number of categories of words available as cues for recall, because the subjects acted as their own controls. Nor is it likely that they can be explained in terms of the length of time taken to process an item: This variable failed to correlate with memorability in Experiment 2, and other experiments have also shown that it is a poor predictor of memory performance (e.g., Craik & Tulving, 1975).

The majority of subjects in Experiment 1, where there were three components in the target category, reported that they had carried out the categorization task in a serial fashion. Whether processing is really serial is impossible to determine, but it is interesting to note that the recall data are consistent with this assumption. If a subject selects one of the target components at random and tests whether the meaning of a given word contains it, then it is easy to show that the mean number of tests required for the four sorts

of words are as follows: 1 test for a word with none of the target components, 1.33 tests for a word with one target component, 2.0 tests for a word with two target components, and 3.00 tests for a word with three target components. The results of Experiment 1 are in accord with these figures, particularly the slight difference in recall between words with none of the target components and words with one of them.

The fact that the meaningful processing of words need not necessarily involve all of their semantic components naturally leads to the further question of what is retrieved about them in the ordinary comprehension of discourse. One recent conjecture is that initially no direct semantic information is recovered (Fodor, Fodor, & Garrett, 1975). However, there are grounds for supposing that more is involved even in the initial stages of comprehension (Johnson-Laird, 1975). Although it is extremely difficult, if not impossible, to investigate normal comprehension experimentally without disrupting or distorting it, some information about lexical items is likely to be retrieved: What information probably depends on the context in which they occur. Indeed, Hodgkin (1977) has found that different semantic components of a noun become salient depending upon the verb that occurs in the sentence.

Our assumption that the amount of processing of an item determines its memorability is obviously similar to Craik and Lockhart's (1972) proposal that the "depth" of processing of an item is the crucial factor. These authors assume, for instance, that semantic tasks generally require a deeper level of processing than phonological tasks, and hence lead to a better recall of the experimental materials. Unfortunately, it has so far proved impossible to provide a theoretical rationale for the relative depths of different sorts of tasks: To rely merely on the results of memory experiments is to create a vicious circle (Nelson, 1977). However, the concept of amount of processing can be defined in terms of the number of decisions about an item that yield pertinent information to the task in hand. Although this concept does not distinguish between levels of processing, it predicts without circularity the results of the present experiments. It also accounts for those findings that led the proponents of the depth-of-processing framework to introduce a new explanatory notion, the "richness" or "elaboration" of a memory representation (Craik & Tulving, 1975). It has been shown experimentally that items responded to positively are better remembered than items responded to negatively; for example, when subjects are asked "Is a shark a type of fish?" and "Is a heaven a type of fish?", they remember "shark" better than "heaven" (cf. Schulman, 1974). The notion of "elaboration" suggests that where a positive response is made, the encoding question and the target word form a more coherent and integrated unit. This may well be true, but the amount of elaboration seems hard to assess. We accordingly propose to explicate the notion

in terms of the amount of relevant processing. Such a step has the advantage of leading to more explicit predictions, including differential predictions about the fate of nontarget items. Deciding that a whale is not a fish, for example, would require a greater amount of semantic processing than deciding that a heaven is not a fish: Merely recollecting that a whale swims in the sea does not suffice, one must recall that it has no gills and must surface to breathe, and so on, whereas any aspect of heaven suffices to reject it. Hence, subjects should remember "whale" better than "heaven" after they have rejected both of them as varieties of fish. It is unclear what prediction, if any, the notion of "elaboration" would make here.

In conclusion, the present experiments suggest that some semantic components of a word can be retrieved and processed without having to process all of them, and that the greater the amount of semantic processing, the more likely the word is to be remembered. The concept of amount of processing can, of course, be extended to deal with tasks at any level from phonology to pragmatics.

REFERENCES

- COLLINS, A. M. & QUILLIAN, M. L. How to make a language user. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972.
- CRAIK, F. I. M., & LOCKHART, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 671-684.
- CRAIK, F. I. M., & TULVING, E. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 1975, 104, 268-294.
- FODOR, J. D., FODOR, J. A., & GARRETT, M. F. The psychological unreality of semantic representations. *Linguistic Inquiry*, 1975, 4, 515-531.
- FRASE, L. T., & KAMMANN, R. Effects of search criterion upon unanticipated free recall of categorically related words. *Memory & Cognition*, 1974, 2, 181-184.
- GIBSON, E. J. Perceptual learning and the theory of word perception. *Cognitive Psychology*, 1971, 2, 351-368.
- HODGKIN, D. *An experimental study of sentence comprehension and semantic memory*. Unpublished PhD dissertation, University of London, 1977.
- HYDE, T. S. Differential effects of effort and type of orienting task on recall and organization of highly associated words. *Journal of Experimental Psychology*, 1973, 97, 111-113.
- HYDE, T. S., & JENKINS, J. J. Differential effects of incidental tasks on the organization and recall of a list of highly associated words. *Journal of Experimental Psychology*, 1969, 82, 472-481.
- JOHNSON-LAIRD, P. N. Meaning and the mental lexicon. In A. Kennedy & A. Wilkes (Eds.), *Studies in long term memory*. London: Wiley, 1975.
- KLEIN, K., & SALTZ, E. Specifying the mechanisms in a levels-of-processing approach to memory. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, 2, 671-679.
- KUCERA, H., & FRANCIS, W. N. *Computational analysis of present-day American English*. Providence, RI: Brown University Press, 1967.
- MANDLER, G. Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation*. New York: Academic Press, 1967.
- MILLER, G. A., & JOHNSON-LAIRD, P. N. *Language and perception*. Cambridge, Mass: Harvard University Press and Cambridge: Cambridge University Press, 1976.
- NELSON, T. O. Repetition and depth of processing. *Journal of Verbal Learning and Verbal Behavior*, 1977, 16, 151-171.
- ORNSTEIN, P., TRABASSO, T., & JOHNSON-LAIRD, P. N. The effects of instructions to organize and to recall. *Journal of Experimental Psychology*, 1974, 16, 1014-1018.
- RIPS, L. J., SHOEN, E. J., & SMITH, E. E. Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 1-20.
- SCHANK, R. C. Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology*, 1972, 3, 552-631.
- SCHULMAN, A. I. Memory for words recently classified. *Memory & Cognition*, 1974, 2, 47-52.

(Received for publication January 26, 1978;
accepted March 28, 1978.)