# The effects of search criteria and retrieval cue availability on memory for words

A. G. R. McCLELLAND, R. E. RAWLES, and F. E. SINCLAIR
*University College London, London WC1E 6BT, England*

The effects of search criteria upon recall performance following a word classification task were investigated. Rival hypotheses were tested, one based on the number of cues available at retrieval and the other on the amount of processing carried out during classification. After being presented with two-component category names, subjects classified words on the basis of either category membership or the number of semantic components the words shared with a target category, and then they received an unexpected recall test. Experiment 1 favored an amount-of-processing explanation, but it was felt that redundancy in the stimulus materials might invalidate this conclusion. Experiment 2 confirmed this suspicion and clearly supported an explanation in terms of retrieval cue availability. The notion of partial semantic processing of words is discussed, and it is argued that the number of congruent encodings formed during classification can be used as an index of elaboration and subsequent memory performance.

Johnson-Laird, Gibbs, and de Mowbray (1978) recently reported the results of two experiments which suggested that some elements of the meaning of a word can be processed without having to process them all. They manipulated the amount of processing required to classify a list of words as members or nonmembers of a given target category and obtained improved recall with increased processing.

The results were of great interest, not only because they appeared to resolve the controversy surrounding the retrieval of semantic components of words (i.e., are all or only some of the components of a word retrieved when it is encountered in a sentence), but also because the experimental manipulation promised an independent measure of elaboration (Craik & Tulving, 1975) in terms of the number of classificatory decisions made about a word that yield pertinent information. As Johnson-Laird et al. (1978) point out, this has the advantage of leading to more explicit predictions concerning the fate of nontarget items in a classification experiment. For example, in their second experiment, they employed two-component categories, such as "consumable solid." On the basis of their argument, it should take a subject 2.0 decisions to classify a target item (e.g., bread), 1.5 decisions, on average (assuming random access of the components), to classify a word with one component in common with the target category (e.g., sherry, iron), and 1.0 decisions to classify an item with neither of the components (e.g., shampoo). The last two types of words are both nontargets, but

differential recall was predicted and obtained, based on the different number of decisions required during classification.

In their paper, Johnson-Laird et al. (1978) specifically dismissed an explanation of their results in terms of retrieval cues because, unlike some previous studies (e.g., Frase & Kammann, 1974; Klein & Saltz, 1976), they held the number and type of word categories constant across encoding conditions, thus ruling out an account in terms of differences in the number or the nature of the retrieval cues available. However, a retrieval explanation is possible if one considers the category names' components rather than the category names per se as being the most likely retrieval cues. Thus with a two-component category name, there are potentially two cues available for target items, one cue for words with one component, and no obvious cue for words with neither of the target components.

In order to test this alternative hypothesis, it is necessary to vary the number of decisions made about a word during classification independently of the number of components it has in common with the target category. In the situation in which subjects have to search for target items (target search), as in Johnson-Laird et al.'s (1978) experiments, these two factors are always confounded. To achieve the required independence, a new task was employed in the present study in addition to target search. This new search criterion (component search) consisted of asking subjects to indicate for each word in the list the number of semantic components the word had in common with the target category. With a two-component category name, the subjects must make two decisions about each word in the list, irrespective of whether it has two, one, or none of the components in common with the target category.

The first experiment was a partial replication of

Johnson-Laird et al.'s (1978) Experiment 2, employing component-search instructions in addition to the target-search instructions they had used. Clearly, for the target-search condition, the same results obtained by Johnson-Laird et al. would be predicted on the basis of either an "amount-of-processing" or a "number of retrieval cues" hypothesis. However, for the component-search conditions, there are two distinct alternatives. Amount of processing would predict no difference in recall across the three component levels and a consequent interaction between search strategy (target or component) and component number. An explanation in terms of the number of retrieval cues available would predict a pattern of results similar to those for target search and, consequently, no interaction.

## EXPERIMENT 1

### Method

The subjects were run in two groups, one group receiving target-search and the other receiving component-search classification instructions. Within these groups, subjects were assigned at random to one of the four following target categories: consumable solids, consumable liquids, nonconsumable solids, and nonconsumable liquids. Each subject searched through the same word list, although the order of presentation was randomized across subjects. Following the classification task, an unexpected free recall test was administered.

**Materials.** The stimulus list consisted of 10 words from each of the four categories (bread, cake, garlic, peanut, pear, radish, rice, salmon, steak, toffee; coffee, gravy, lager, lemonade, milk, sherry, soup, syrup, vinegar, vodka; bronze, concrete, gravel, jade, nylon, plastic, ruby, silk, wool, zinc; acid, dye, glue, ink, lotion, paint, paraffin, perfume, shampoo, varnish). As far as possible, the words were balanced across categories for length and frequency of usage (Kučera & Francis, 1967). The list was presented in a booklet consisting of a front cover, a page with the target category printed on it, and four pages of words, each word printed in uppercase and placed on a separate line. The words were assigned at random to the pages, and the order of pages in the booklets was also randomized.

**Subjects.** Thirty-two volunteer subjects participated in the experiment. They were mostly junior and senior high school students, together with some 1st-year undergraduates from the department of psychology, University College London. None of the subjects had previously taken part in an experiment of this kind.

**Procedure.** Subjects were assigned at random to one of the experimental conditions and were run under either target- or component-search instructions, respectively. Both groups were informed that the experiment was concerned with words and how people classified them into categories. Subjects were told that the booklets in front of them consisted of a page giving the name of a target category followed by four pages of words. Under target-search instructions, they were asked to read the category name carefully and to proceed through the booklet putting a checkmark by category members and a cross by non-members. Under component-search instructions, an example of each type of category/item pairing was given. The category used was "dangerous fish," and the two-, one-, and zero-component words presented as examples were "SHARK," "POISON," "TROUT," and "PAPER," respectively. The subjects were instructed to go through the booklet putting a 2, 1, or 0 by each word as an indication of how many components that word had in common with the target category. Both groups were asked to work as quickly and accurately as possible and to close their booklets as soon as they had finished. All subjects completed the task within 2 min and were instructed to turn their booklets over and write down on the back cover as many

of the words from the list as they could. Three minutes were given for recall.

### Results

The proportions of words recalled at the three component levels and the overall recall under the two search conditions are shown in Table 1. The particular target category searched for did not have a reliable effect on recall performance [F(3,24) = 1.06]. Recall following component search was significantly better than that following target search [F(1,24) = 6.31, p < .025]. The number of semantic components in common with the target category reliably affected recall [F(2,48) = 18.14, p < .001], and this factor interacted with both search task [F(2,48) = 10.37, p < .001] and category type [F(6,48) = 3.86, p < .05]. Planned comparisons showed that under target-search conditions, component number was a reliable effect [F(2,24) = 33.72, p < .001], but under component-search conditions, no such effect was found [F(2,24) = .46].

Further analysis revealed a significant trend across components in the target-search condition (Page's L = 218, p < .001), but no trend in the component-search condition (Page's L = 192).

### Discussion

The results indicate that amount of processing significantly affects recall and that the number of decisions made about a word can be used as an index of elaboration. With target-search instructions, the results replicate those obtained by Johnson-Laird et al. (1978, Experiment 2), whereas with component-search instructions, no differences were found in recall performance across the three component levels. Thus, it appears that it is the number of decisions made during encoding rather than the number of congruent components available as retrieval cues that is the crucial variable affecting recall performance.

However, in comparison with results from other experiments, the present data are rather surprising. For example, Craik and Tulving (1975) noted that in the classification tasks they used, words that were responded to positively (i.e., category members) were generally better recalled and recognized than were words responded to negatively. Indeed, they introduced the notion of elaboration in order to explain this result, arguing that positive instances could be encoded in an integrated fashion with the encoding question because they were "congruous" with it (Schulman, 1974), whereas negative

**Table 1**
**Proportions of Words Correctly Recalled in Experiment 1**

| | Number of Target Category Components Possessed by the Words | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | Two | | One | | Zero | | Overall | |
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| Target Search | .48 | .13 | .27 | .10 | .16 | .12 | .29 | .16 |
| Component Search | .39 | .15 | .36 | .12 | .35 | .15 | .37 | .20 |

instances could not, this difference leading to the category member advantage.

It is clear that in the present experiment, negative instances are not congruent with the target category and hence these category name/item encodings should not be well integrated, and, consequently, the items should not be well recalled. In fact, under component-search instructions, all types of item were equally well recalled.

A possible explanation for the discrepancy between the present results and those obtained by Craik and Tulving (1975) lies in the stimulus materials used and the natures of the classification tasks employed. In the present study, a subject in the component-search condition was forced to consider both components of a word and may have classified congruently with regard to nontarget components. By deciding that a given item does not have the component "consumable," a subject might note that it is, by default, "nonconsumable." Similarly, an item that is not "solid" is "liquid" (there being no other phase states in the list). Thus, given that the subject could remember the four components, recall could again be explained in terms of available retrieval cues, each item having been encoded congruently with two of the four components.

In Craik and Tulving's (1975) experiments, not only were more than four categories used, but also the categories were semantically unrelated and were embedded in a presentation sequence containing, in addition, non-semantic (rhyme and case) questions. Under these circumstances, first, subjects would be far less likely to reclassify nontarget stimuli in terms of their positive components, and second, subjects should find that generating these components as retrieval cues would become difficult. Thus Craik and Tulving's (1975) stimuli lacked the redundancy found in the materials used in Experiment 1.

To investigate whether the differences between the studies outlined above could account for the contradictory results obtained, an experiment was carried out using 16 semantically independent categories, rather than the 4 semantically related categories used in Experiment 1. Again, target- and component-search tasks were employed, but in addition to free recall, a cued recall measure of retention was taken.

It was predicted that the pattern of results would be the same for both target- and component-search conditions, with recall performance on both measures declining as the number of retrieval cues available was reduced. With the assumption that recall cues are used under free as well as cued recall conditions, in the latter case zero-component words should be unaffected by presentation of cues, but as more potential cues are made available, one-component words should show a recall advantage and two-component words a greater advantage still. This leads to the prediction of an interaction between recall measure and the number of components in common with the target category.

## EXPERIMENT 2

### Method

The design of the experiment followed the same pattern as Experiment 1, with only minor modifications to incorporate the use of multiple categories and cued recall.

**Materials.** The stimuli consisted of 16 two-component category names, each category appearing four times in the list. Each category appeared once with a two-component item, twice with a one-component item, and once with a zero-component item, giving a total of 64 words for classification. (For category names and list items, see the Appendix.) The order of the 16 category names was randomly assigned to the booklet pages, each page of the booklet containing 4 categories. The order of the first 4 pages in any booklet determined the order of the remaining 12 such that the lag between repetitions of a given category was constant. The first four pages were presented in all possible orderings, giving 24 booklets in all. Thus each subject received the same category names and words for classification, but each booklet presented the material in a unique order.

Subjects were also given a second booklet for their cued recall responses, which consisted of 16 pages, one category name to a page. The order of the categories for any subject was the same as the order in his/her first booklet, so, again, the lag between presentations remained constant.

**Subjects.** There were 48 subjects, mostly junior and senior high school students, together with some undergraduates from University College London.

**Procedure.** The procedure was the same as that used in Experiment 1, except that the instructions were modified to take into account the multiple categories and the cued recall phase of the experiment. In case they thought that they had received defective booklets, subjects were warned that the categories were repeated. The different demands of the two classification tasks were explained as in Experiment 1. Free recall instructions were as before, but 4 min were given for recall. Subjects were then asked to turn to their second (cued recall) booklets. The experimenter explained that on each page they would find one of the category names from the previous booklet and that their task was to try to write down the four words that had been paired with it. This task was paced by the experimenter at a rate of 15 sec/page. This procedure also took 4 min to complete.

### Results

The proportions of words recalled are shown in Table 2. The classification task (target vs. component) did not have a reliable effect on recall $[F(1,46) = 2.90]$. Cued recall was superior to free recall $[F(1,46) = 126.39, p < .0001]$. The number of semantic components in common with the target category was a significant factor $[F(2,92) = 316.79, p < .0001]$, and this interacted with both search task $[F(2,92) = 18.41, p < .0001]$ and recall measure $[F(2,92) = 170.93, p < .0001]$. The three-way interaction among search task, recall measure, and component level was also significant $[F(2,92) = 3.71, p = .028]$.

In addition, recall proportions under the two search conditions were analyzed separately.

**Target search.** Cued recall was superior to free recall $[F(1,23) = 52.48, p < .0001]$. Also, the number of components in common with the target category was a reliable effect $[F(2,46) = 283.17, p < .0001]$, as was the interaction between these two factors $[F(2,46) = 78.38,$

Table 2
Proportions of Words Correctly Recalled in Experiment 2

| | Number of Target Category Components Possessed by the Words | | | | | | | | | | | |
| | Free Recall | | | | | | Cued Recall | | | | | |
| | Two | | One | | Zero | | Two | | One | | Zero | |
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| Target Search | .56 | .09 | .33 | .11 | .21 | .09 | .90 | .11 | .42 | .14 | .20 | .10 |
| Component Search | .38 | .15 | .30 | .10 | .25 | .10 | .75 | .17 | .47 | .17 | .21 | .14 |

p < .0001]. There was a significant trend across components both within the free recall (Page's L = 332, p < .001) and within the cued recall (Page's L = 335, p < .001) scores.

**Component search.** As predicted, the results were very similar to those obtained under target-search conditions, with main effects of recall measure [F(1,23) = 74.53, p < .0001] and components [F(2,46) = 85.77, p < .0001] and a significant interaction between them [F(2,46) = 96.21, p < .0001]. Again, there was a significant trend across components both under free recall (Page's L = 313, p < .001) and under cued recall (Page's L = 334, p < .001) retrieval conditions.

## GENERAL DISCUSSION

In contrast to Experiment 1, the general pattern of increasing recall with an increasing number of congruent components was found in Experiment 2 under target-search and component-search instructions. This was true for both free and cued recall, and the predicted interaction between recall measure and number of components was obtained.

These data lend support to the notion that subjects in the first experiment had indeed used nontarget components as retrieval cues; they are in agreement with the results obtained by Craik and Tulving (1975) and cast doubt upon Johnson-Laird et al.'s (1978) explanation of their own results in terms of amount of processing.

It is proposed that the critical variable affecting memory performance is the number of congruent components available at retrieval. Thus the number of decisions leading to congruent encodings rather than the total number of decisions should be used as an index of elaboration. Processing that does not lead to the formation of congruent encodings will not enhance recall, even if this processing is pertinent to the classification task. In principle, there is no reason why this measure should be restricted to semantic encodings and, presumably, congruent nonsemantic encodings should similarly enhance recall performance.

Additional evidence to support the view that amount of processing per se does not lead to improved memory is given by Stein, Morris, and Bransford (1978). Using words embedded in sentences as stimuli and sentence frames (i.e., the sentences minus the target words) as cues, they showed that only when the sentence was elaborated with information congruent with the target item did recall for the item improve. Indeed, they found that noncongruent elaboration of a sentence led to a decrease in recall performance.

The results of the present investigation bring into question Johnson-Laird et al.'s (1978) assertion that subjects are capable of processing only some of the semantic attributes of a word during reading. This conclusion was perhaps already questionable, based as it was on the assumption that the processing required to classify a word in their experiments would be the same as that required to read the word when it was encountered in a sentence. It is clear that the task demands in word classification experiments such as those employed in the present study and by Johnson-Laird et al. (1978) directly affect subsequent memory for the material classified. This does not, however, rule out the possibility that a two-stage process operates in such experiments. The subjects may first read the word (activating all or some of the semantic components) and then proceed to classify the stimulus on the basis of those attributes relevant to the task.

There is certainly a good deal of evidence to suggest that subjects either consciously or unconsciously access attributes of words not required in order to perform an orienting task. Klein (1964), in his examination of the processes underlying the Stroop (1935) phenomenon, showed that subjects found it impossible not to extract semantic information from a word when their task was to name the color of ink in which it was printed. Jacoby and Craik (1979) noted that the nonzero level of retention following a nonsemantic judgment such as "Is the word in uppercase?" provides evidence that information beyond that required to perform the task must have been accessed. Nelson (1979) gives details of several experiments that support this view. Thus the controversy surrounding the extent to which a word is semantically analyzed in the context of a sentence remains unresolved.

The interaction between component level and recall measure lends support to the view that retrieval cues are used under free recall as well as under cued recall conditions. Particularly strong evidence is provided by the fact that when retrieval cues were presented for zero-component words, recall was no better than it was under free recall conditions. With one-component words, recall was markedly better when the cues were

provided, and recall was virtually doubled for the two-component words when cues were given.

Also noteworthy in Experiment 2 is the interaction between the number of semantic components in common with the target category and the search task. Examination of the recall proportions under target-search conditions reveals that in comparison with one- and zero-component words, two-component items are associated with higher scores than would be expected solely on the basis of the number of encodings formed during classification. Under component-search conditions, in which two-component items were not defined as "target items," the advantage disappears, and thus it is highly likely that this gave rise to the interaction. Johnson-Laird et al. (1978) suggested that such an advantage might exist (and it can clearly be seen in their results, as well as in the target-search data from Experiment 1), but they did not propose an explanation. The present findings indicate that the effect is critically dependent upon the wordings of the search instructions given to subjects, but at the moment, its exact cause remains obscure.

In conclusion, the present study suggests that the number of decisions leading to congruent encodings in a word classification task can be used as an efficient measure of elaboration. Further, the results bring into doubt the assertion made by Johnson-Laird et al. (1978) that some components of a word can be retrieved and processed without having to process them all, and it is suggested that the verdict on partial semantic activation of a word remains open. Finally, the results from the study are in line with the more general theoretical viewpoint (e.g., Jones, 1979; Tulving, 1979) that recall occurs only as a consequence of a cue being present, either externally as in the case of cued recall or internally as a mental event in the case of free recall.

## REFERENCES

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.

FRASE, L. T., & KAMMANN, R. Effects of search criterion upon unanticipated free recall of categorically related words. *Memory & Cognition*, 1974, 2, 181-184.

JACOBY, L. L., & CRAIK, F. I. M. Effects of elaboration of processing at encoding and retrieval: Trace distinctiveness and recovery of initial context. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory*. Hillsdale, N.J: Erlbaum, 1979.

JOHNSON-LAIRD, P. N., GIBBS, G., & DEMOWBRAY, J. Meaning, amount of processing, and memory for words. *Memory & Cognition*, 1978, 6, 372-375.

JONES, G. V. Analyzing memory by cuing: Intrinsic and extrinsic knowledge. In N. S. Sutherland (Ed.), *Tutorial essays in psychology: A guide to recent advances*. Hillsdale, N.J: Erlbaum, 1979.

KLEIN, G. S. Semantic power measured through the interference of words with color-naming. *American Journal of Psychology*, 1964, 77, 576-588.

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.

KUČERA, H., & FRANCIS, W. N. *Computational analysis of present-day American English*. Providence, R.I: Brown University Press, 1967.

NELSON, D. L. Remembering pictures and words: Appearance, significance and name. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory*. Hillsdale, N.J: Erlbaum, 1979.

SCHULMAN, A. I. Memory for words recently classified. *Memory & Cognition*, 1974, 2, 47-52.

STEIN, B. S., MORRIS, C. D., & BRANSFORD, J. D. Constraints on effective elaboration. *Journal of Verbal Learning and Verbal Behavior*, 1978, 17, 707-714.

STROOP, J. R. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 1935, 18, 643-662.

TULVING, E. Relation between encoding specificity and levels of processing. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory*. Hillsdale, N.J: Erlbaum, 1979.

## Appendix
### Category Names and Associated List Items Employed in Experiment 2

| | Number of Components in Common with the Category Name | | | |
| --- | --- | --- | --- | --- |
| | Two | One | One | Zero |
| 1. American Author | Hemingway | Baseball | Dickens | Cricket |
| 2. Consumable Solid | Bread | Sherry | Brass | Shampoo |
| 3. European Country | Belgium | Paris | Canada | New York |
| 4. Fast Transport | Airliner | Cheetah | Donkey | Snail |
| 5. Heavy Furniture | Wardrobe | Boulder | Chair | Pebble |
| 6. Human Dwelling | Cottage | Child | Kennel | Spider |
| 7. Indoor Sport | Squash | Kitchen | Rugby | River |
| 8. Liquid Fuel | Paraffin | Milk | Charcoal | Plastic |
| 9. Man-Made Fiber | Polyester | Sausage | Cotton | Honey |
| 10. Marine Mammal | Porpoise | Cruiser | Monkey | Bicycle |
| 11. Military Title | Major | Rifle | Countess | Pencil |
| 12. Musical Instrument | Viola | Choir | Scalpel | Crowd |
| 13. Outdoor Profession | Forester | Meadow | Lawyer | Fireplace |
| 14. Precious Stone | Emerald | Platinum | Granite | Steel |
| 15. Religious Building | Mosque | Hymn | Barn | Ballad |
| 16. Wild Bird | Sparrow | Wolf | Chicken | Cow |