

# Classifying pictures and words: Implications for the dual-coding hypothesis

PAUL L. HARRIS, PETER E. MORRIS, and ELIZABETH BASSETT  
*University of Lancaster, Lancaster, LA1 4YF, England*

In deciding that two items come from the same category, subjects were delayed if the items were presented as a word and a picture, as compared with two words or two pictures. Subjects were not delayed if the items were not to be compared. Hence, the delay appears to occur during the retrieval and comparison of knowledge about the items, not during their perceptual analysis. The results suggest that there are two distinct codes involved in the categorization of words and pictures, rather than a single code.

When presented with either a picture of a flower or the word "flower," one is able to classify both as instances of a more general category, such as "plant." The extent to which subjects exploit the categorical nature of item lists in free recall experiments using pictures or words suggests that category information is both freely available and of importance in cognitive processing (Paivio, 1971). Are pictures and words categorized by identical processes? Categorization might be based on a set of features common to an abstract representation of both the word and the picture. Such an abstract "code," neutral as between pictures and words, has been postulated by Chase and Clark (1972) as a way of matching pictures and sentences.

This account predicts that it should take as long to categorize a word as to categorize a picture, assuming that both can be converted as rapidly into the hypothetical neutral code. Alternative accounts, however, are plausible. For example, it might be assumed that classification is primarily a verbal task. This point of view has a long history in developmental (Vygotsky, 1962), cross-cultural (Bruner, Greenfield, & Olver, 1966), and clinical psychology (Goldstein, 1959). It implies that superordinate categories, such as "plant" are retrieved via some abstract verbal representation of the word "flower." In this case, a picture might have to be named and/or converted into such an abstract verbal code before categorization.

A criticism of the verbal-code hypothesis is that children can categorize some objects before they can name them (Rosch, 1976). This criticism might be taken as support for the neutral-code hypothesis, or for a third alternative, which is to argue that meaning, particularly for concrete nouns, is stored predominantly in terms of images or rules for the regeneration of images. The possibility that concrete sentences are

stored as images has been suggested by Begg and Paivio (1969). If such imagery were retrieved more readily from pictures than from words, then pictures would be categorized more rapidly.

A fourth possibility is that there are separate verbal and visual codes (Paivio, 1971, 1975). The classification of a word will make use of information stored with the verbal code, while the classification of pictures uses separately stored information relevant to visual analyses. If subjects can decide that two words or two pictures belong to the same category more quickly than when the comparison is between a word and a picture, models of the fourth type could most easily account for the difference.

## EXPERIMENT 1

Experiment 1 was conducted as a first step toward distinguishing between the above possibilities. Subjects were presented with either two names of physical objects, or two pictures of objects, or a name and a picture. They were asked to report whether the two items came from the same or different categories. No difference in latencies for the three conditions would support a neutral-code hypothesis. Quicker responses for word-word comparisons would suggest that coding was basically verbal, while quicker picture-picture comparisons would implicate imagery coding. If word-picture comparisons proved the longest, then a dual code would be suggested.

### Method

**Subjects.** Fifteen undergraduates from the psychology department volunteered to participate.

**Procedure and apparatus.** Subjects were presented with a pair of stimuli in a tachistoscope for a period of 2 sec. The onset of the stimulus triggered a millisecond timer which halted when the subject pressed either a "same" or a "different" button, indicating that the two stimulus items came from the same or different categories. The categories used were "animals" (with pig, dog, cat, and rat as the items) and "modes of transportation" (with car, bicycle, train, and boat as items).

Paul Harris is now at the Department of Psychology, Free University of Amsterdam, Netherlands.

Each item was used once in each of the four combinations of word-picture and same-different categories. There were 32 stimulus pairs. Eight of the pairs contained two words (WW); eight contained two pictures in the form of line drawings, depicting each of the words (PP). The remaining 16 cards contained both a picture and a word; eight cards had the word on the left of the fixation point and the picture to the right (WP) and eight had the reverse arrangement (PW). Within each of these subgroups, half of the pairs required a "same" judgment and half required a "different" judgment. For each subject, the set of 32 cards was presented twice in a different random order. Items on which subjects made an error were rerun, by placing the item randomly in the pile of cards yet to be presented.

**Results**

Separate analyses were carried out for conditions requiring "same" and "different" judgments. Within these conditions, separate analyses were performed for the first and the second presentation of each item, in case changes in performance occurred through experience with the items. For each analysis, there were two random factors, subjects (15) and items (4), and one fixed factor, treatments (4). This required the calculation of quasi F (F') ratios to test the treatment effects (Clark, 1973). For "same" judgments, first presentation, there was a significant treatments effect [F'(5,33) = 3.89, p < .01, MSe = 861].<sup>1</sup> Three orthogonal comparisons between the treatments were conducted. There was no difference between the WW and PP conditions [F(1,33) = 1.00], nor between the WP and PW conditions [F(1,33) = .02]. However, the WW and PP conditions were significantly faster than the PW and WP conditions [F(1,33) = 8.50, p < .01].

For "same" judgments, on the second presentation of items, a similar pattern of results was found. There was a significant treatment effect [F'(4,17) = 3.65, p < .05, MSe = 768]. Neither the comparison of WW and PP [F(1,33) = .04], nor of WP and PW [F(1,33) = 2.17] were significant, but the WW and PP conditions were significantly faster than the WP and PW conditions [F(1,33) = 7.65, p < .01].

For "different" judgments, there was no significant effect of treatments on either the first [F'(9,26) = .84, MSe = 1,068] or the second presentation [F'(7,32) = 1.16, MSe = 545]. The mean latencies combined for first and second presentations are shown in Table 1.

**Discussion**

Subjects judged two pictures or two words to be the same category in approximately 730 msec. A further

100 msec was required by subjects if the two items were different, either in category or in code. A dual-code hypothesis seems to be implicated by this result. Dual-code models can easily encompass a difference between word-picture and picture-picture or word-word comparisons. A more precise model needs to be developed, but this will be considered later. The results are not compatible with the first two positions described in the introduction. Two pictures were judged as rapidly as two words. Categories, therefore, appear to be neither predominantly word nor imagery based.

It could be argued that the results do not rule out a single-code hypothesis, since access to such a code (verbal, imagery-based, or "neutral") is as fast via a picture as via a word. But in this case, picture and word combinations should be classified as quickly as two words or two pictures. One way of rescuing a single-code hypothesis is to argue that picture-word combinations require a time-consuming switch in the mode of perceptual analysis. At the comparison level, however, pictures and words might be classified in terms of a single code. There is, however, an objection to this hypothesis.

If the analysis of a word-picture involves a switch from a verbal to a visual analyzer, which takes a significant amount of time, so accounting for the difference between conditions for "same" responses, then the "different" judgments in the WP and PW conditions should be longer by a similar amount. Whatever processing occurs must follow the initial perception, and, if that perception is delayed, so will be the response. However, code did not influence the "different" responses, suggesting a similar process was involved in "different" judgments in all four conditions.

Nonetheless, the single-code hypotheses have the advantage of parsimony and should not be rejected too quickly. As a stronger test between the single-code and the dual-code hypotheses, Experiment 2 was conducted.

**EXPERIMENT 2**

The single-code explanation of the difference between PP, WW and PW, WP conditions in Experiment 1 is that there is a switch between a verbal and a visual analyzer in the PW and WP conditions. The time required for switching accounts for the time differences. A dual-code model predicts that the switch is not the source of the delay. A delay will occur only when information about the items is being found and analyzed. A switch between pictures and words without this requirement will not affect latencies.

Subjects were shown two pairs of items at a time. They were asked to report whether the two pairs were of items from the same class. The pairs of items, which could be pictures or words, were presented one above the other. Subjects compared the top two items and the bottom two items, but were not required to make

**Table 1**

Mean Reaction Time (in Milliseconds) for "Same" and "Different" Judgments According to Code in Experiment 1

Code	WW	PP	WP	PW
"Same" Judgments	720	742	855	815
SD	60	69	87	81
"Different Judgments	840	806	855	816
SD	80	74	82	83

any comparison between the upper and the lower items. They responded "same" if both top and bottom pairs matched for category. There were three classes of stimuli: (1) all four items with the same representation (all pictures or all words), (2) the upper pair words, the lower pictures, or vice versa, and (3) both pairs containing a picture and a word. This allowed the comparison of latencies in the three situations. When all the items were words, or all were pictures, there should be no switching of analyzers, according to the single-code hypothesis. According to the dual-code hypothesis, this condition involves the use of only one code. Therefore, according to both theories, the condition involves the use of only one code, and, according to both theories, this condition should produce the fastest responses. Most interesting for discriminating between the theories is the condition in which the top pair of items have one form of representation, while the bottom pair have the other. In that case, there is a switch between the analysis of pictures and words. According to the single-code hypothesis, this switch should take time and increase the latency. However, according to the dual-code hypothesis, in this situation, in contrast to Experiment 1, the switch involves no comparison of items with different codes. The dual-code hypothesis, therefore, predicts no difference in latency between this and the first condition. The third situation is one in which the items to be compared for category differ in form of representation. In that case, comparison involves at least two switches between codes during the analysis for category and both models predict a longer latency than in the first condition. This situation, therefore, provides a replication of Experiment 1.

### Method

**Subjects.** Eighteen undergraduate psychology students volunteered to take part in this experiment.

**Procedure and apparatus.** The method of presentation and testing was the same as in Experiment 1. Items were taken from three categories: "animals" (deer, rat, pig, dog), "clothing" (coat, trousers, shoe, hat), and "modes of transportation" (train, boat, car, bicycle). Seventy-two stimulus cards were prepared, representing six subgroups of 12 cards, namely: (1) all items words, (2) all items pictures, (3) upper pair pictures, lower pair words, (4) upper pair words, lower pair pictures, (5) a picture and a word in each pair, both words on the left, and (6) a picture and a word in each pair, both words on the right. Within each subgroup, half the cards had both pairs with items from the same category (although the pairs came from different categories), thus requiring a "same" response, and half had one pair with items from different categories. The use of the category items was balanced across the groups.

### Results

For each subject, the latency data was combined for Subgroups 1 and 2, 3 and 4, and 5 and 6, yielding latency data for the three relevant conditions, which will be referred to as "same code" (SC), "between-pairs shift" (BPS), and "within-pairs shift" (WPS). Responses in the "same" and "different" conditions

**Table 2**  
Mean Reaction Time (in Milliseconds) for "Same" and "Different" Judgments According to Conditions in Experiment 2

Condition	SC	BPS	WPS
"Same" Judgments	1328	1330	1424
SD	260	300	300
"Different" Judgments	1296	1344	1288
SD	340	340	350

were analyzed separately. The means and standard deviations for the three conditions are given in Table 2.

Inspection of Table 2 reveals that, for "same" judgments, the latencies for the SC and BPS differ by only 2 msec, while the latency in the WPS condition was over 90-msec longer. An analysis of variance, with conditions (3), items (6), and subjects (18) as factors, revealed a significant overall effect for conditions [ $F'(2,42) = 5.77$ ,  $p < .01$ ,  $MSe = 635$ ], which must result from the difference between the WPS and the other two conditions. An analysis for "different" responses showed no significant difference between the conditions [ $F'(4,21) = 1.48$ ,  $MSe = 931$ ]. All 18 subjects were quicker in responding to items when the difference occurred in the upper pair, indicating that subjects tended to process the upper pairs first. Mean latency when the difference was in the upper pair was 1,149 msec; mean latency when the difference was in the lower pair was 1,463 msec.

### Discussion

As predicted by the dual-code hypothesis, there was no difference in latency when the shift between modalities occurred between pairs, and the shift did not involve the comparison of the items within pairs. The longer latency when the shift occurred within pairs repeats the finding of the first experiment. The identification of two items as coming from the same category was quicker when the items were in the same form of representation, either as pictures or as words. Combined with the results of Experiment 1, this experiment casts severe doubts on a single-code hypothesis. Such a hypothesis was only saved after Experiment 1 by postulating a time-consuming switching mechanism when perceptual analysis is switched from words to pictures, or vice versa. Such a delay should have occurred for the BPS condition in this experiment, but it did not. A dual-code model seems necessary to accommodate the results. So far, the characteristics of such a model have been left vague. Below, two possible models are described which will account for the findings of the two experiments. The models must accommodate three facts. First, they must predict faster judgments when two items are presented in the same form of representation than when the items are represented differently. Second, they must take into account the similar latencies for all conditions when the items come from different categories. Third, they should

explain the very close similarity between "same" and "different" judgments for PW and WP conditions in Experiment 1.

### GENERAL DISCUSSION

The first model attributes the differences in latencies to the initial retrieval of the stored information relevant to the presented items. It assumes that words belonging to the same domain are stored relatively close together in memory. Similarly, information relating to images belonging to the same domain is stored close to information on other items from the same domain. The distance analogy is one which has been frequently used in network models of semantic memory (e.g., Collins & Quillian, 1969; Rumelhart, Lindsay, & Norman, 1972). Nor is it incompatible, given certain modifications of the metaphor, with attribute models (e.g., Herriot, 1974; Rips, Shoben, & Smith, 1973; Smith, Shoben, & Rips, 1974). Whenever two items are stored close together, they can be more rapidly retrieved or compared than items from different domains. Shifting to a new domain takes a fixed amount of time; this time is invariant whether the shift is to items in the same category, but a different code, or, alternatively, to items in the same code, but different category. This model fits the results obtained. However, it does contain assumptions about the retrieval system which may be questioned. Is the initial retrieval of information on items influenced by their semantic relatedness? Meyer and Schvaneveldt (1971) found that the identification of two letter strings as words was facilitated if the letter strings formed commonly associated words. This suggests a semantic organization in the storage of items. The assumption that is more open to question is that the retrieval of information on unrelated items is the same, whether the items differ in category or in code.

The second model emphasizes the comparison processes. It is suggested that, when two items are presented, two processes begin. The first process involves the calling of the list of attributes appropriate to the items; these attributes are compared for the two items. If the relevant attributes match, then the items will be reported as coming from the same category. The second process is the categorization of the item and the comparison of the categories of the two presented items; a match leading to a "same" response, and a mismatch leading to a "different" response. The two additional assumptions are (1) that the retrieval of an item's attributes is faster than the retrieval of its category (perhaps because categorization itself depends on the recall of attributes); and (2) that items presented visually and verbally lead to the retrieval of different sets of attributes. The latter assumption is supported by the results of Karwoski, Gramlich, and Arnott (1944), who report different associates to pictures of

objects than to the objects' names. This model predicts that, for WW and PP "same" trials, attribute matches should give a quicker response than in the other conditions, when the attribute match will fail and the decision must depend upon the second process. One attraction of this model is that, not only does it account for the similarity of all the "different" responses and the "same" response for PW and WP in Experiment 1, but also, it deals with the very close similarity of "same" and "different" judgments for PW and WP. The latter latencies are so similar that it would be difficult to believe that basically different processes are involved. According to the model, the same process takes place. The items are categorized, and the categories are compared and judged as same or different, on the basis of their match or mismatch. The part of this model dealing with the retrieval and comparison of attributes is similar to the initial stages of the attribute-comparison model of Smith et al. (1974), which provides an adequate account of the comparison processes in semantic memory tasks.

The adequacy and implications of the models will be the subject of future research. What can be concluded from the experiments reported here is that, when pictures and words are analyzed for their meaning, the analysis does not take place in the same neutral code, nor in purely verbal or visual codes. Information relevant to both the verbal and visual representation is involved. For these representations, there are differences either in the organization of stored information, or in the nature of the information stored, and its influence upon the decision processes involved. Such differences occur when the meaningful properties of the items are considered and do not come merely from differences in the perceptual analysis of words and pictures.

### REFERENCES

- BEGG, I., & PAIVIO, A. Concreteness and imagery in sentence meaning. *Journal of Verbal Learning and Verbal Behavior*, 1969, 8, 821-827.
- BRUNER, J. S., GREENFIELD, P. M., & OLVER, R. R. *Studies in cognitive growth*. New York: Wiley, 1966.
- CHASE, W. G., & CLARK, H. H. Mental operations in the comparisons of sentences and pictures. In L. Gregg (Ed.) *Cognition in learning and memory*. New York: Wiley, 1972.
- CLARK, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 335-359.
- COLLINS, A., & QUILLIAN, M. R. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1969, 8, 240-247.
- GOLDSTEIN, K. Concerning the concreteness in schizophrenia. *Journal of Abnormal and Social Psychology*, 1959, 59, 146-148.
- HERRIOT, P. *Attributes of memory*. London: Methuen, 1974.
- KARWOSKI, T. F., GRAMLICH, F. W., & ARNOTT, P. Psychological studies in semantics: 1. Free association reactions to words, drawings and objects. *Journal of Social Psychology*, 1944, 20, 233-247.
- MEYER, D. E., & SCHVANEVELDT, R. W. Facilitation in recog-

- nizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 1971, **90**, 227-234.
- PAIVIO, A. *Imagery and verbal processes*. New York: Holt, Rinehart, & Winston, 1971.
- PAIVIO, A. Perceptual comparisons through the mind's eye. *Memory & Cognition*, 1975, **3**, 635-647.
- 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.
- ROSCH, E. R. Objects in natural categories. *Cognitive Psychology*, 1976, **8**, 382-439.
- RUMELHART, D. E., LINDSAY, P. H., & NORMAN, D. A. A process model for long-term memory. In E. Tulving & W. Donaldson (Eds.) *Organization and memory*. New York: Academic Press, 1972.
- SMITH, E. E., SHOEN, E. J., & RIPS, L. J. Structure and process in semantic memory: A feature model for semantic decisions. *Psychological Review*, 1974, **81**, 214-241.
- VYGOTSKY, L. *Thought and language*. Cambridge: M.I.T., 1962.

NOTE

1. To simplify the analysis, in both experiments, each score was rounded to centiseconds. This procedure leads to a MSE which is smaller than would be expected if the calculations were in milliseconds. The mean values given in the tables are based on the raw millisecond data.

(Received for publication June 12, 1975;  
revision received July 16, 1976.)