Effects of Decision Difficulty on Recognition and Recall

LARRY L. JACOBY McMaster University

FERGUS I. M. CRAIK University of Toronto

AND

IAN BEGG McMaster University

Four experiments explored the idea that differences in the difficulty of semantic judgments affect the subsequent retention of items involved in the judgments. In general, memory was better after more difficult decisions, although there were interactions with the form of the retention test. It is suggested that decision difficulty is positively related to the distinctiveness of the memorial record of the initial event, and that it is distinctiveness, rather than depth or elaboration, of the memory trace that underlies high levels of memorability. The major difference between distinctiveness and depth as explanatory concepts lies in the different views of meaning offered by the two accounts. In particular, the distinctiveness notion leads to a relativistic view of meaning, in that the meaning of an event is determined by the set of contrasts generated during initial processing.

What are the major factors underlying high levels of retention? In the literature on memory for verbal material, a recurring theme is that retention reflects the degree to which the meaning of the material was involved in the initial learning operations. One viewpoint that has stressed the importance of semantic analysis for memory is the levels of processing framework proposed by Craik and Lockhart (1972). However, as Baddelev (1978) has pointed out, the levels of processing approach has had little to say about the differential memorability of items processed within the semantic domain. In the present article we report four experiments illustrating aspects of the notion that an increase in task difficulty, or the necessity of carrying out more extensive processing within the semantic domain, results in higher levels of retention.

This research was supported by Grants A0281, A8261, and A8122 from the National Research Council of Canada. Requests for reprints should be sent to Larry L. Jacoby, Department of Psychology, McMaster University, Hamilton, Ontario L8S 4K1, Canada. Are such differences best conceptualized as reflecting different initial levels of processing, or is some other description preferable?

Other findings are already difficult to fit into the levels of processing framework. For example, Mandler and Worden (1973) reported an experiment in which the processing of meaning was associated with poor subsequent retention. Frase and Kammann (1974) showed that differences in the specificity of classes used to categorize words resulted in different levels of retention, and Klein and Saltz (1976) found that words rated on two dimensions were better retained than words rated on one dimension. In general, the problem is that there is no satisfactory specification of what constitutes a meaningful analysis (Nelson, 1977) and no clear agreement on how degrees of meaningfulness might be stratified within the semantic domain. Alternative approaches have been framed in terms of elaboration of processing (e.g., Craik & Tulving, 1975), extensiveness of processing (Kolers, 1975), and amount of semantic processing (JohnsonLaird & Bethell-Fox, 1978; Johnson-Laird, Gibbs, & de Mowbray, 1978). These accounts are intuitively appealing, yet they also run into the difficulty of specifying elaboration, extensiveness, or amount in all but the most general terms.

The experiments reported in the present article show that difficult initial decisions are associated with high levels of subsequent retention; the theoretical problem is to give an adequate account of this finding. One possibility is that difficult tasks usually take longer to execute, resulting in longer learning times for the material in question. Although it is well established that increases in study time enhance retention, it has also been demonstrated that processing time by itself is not a good predictor of memory performance (Craik & Tulving, 1975; Kolers, 1973; Johnson-Laird et al., 1978). A second possibility is that greater amounts of processing effort underlie high levels of retention. But how is "effort" specified or measured? Psychophysiological indices are available (Kahneman, 1973) but tell us little about the nature of the concurrent mental operations. It seems strongly preferable to couch our theoretical descriptions in terms of the cognitive operations themselves.

It may well be possible to interpret effects of decision difficulty in terms of differences in meaningfulness or level of processing. However, to do so requires a view of meaning that differs from that implicitly held by the original levels of processing view. If differences in retention are to be interpreted in terms of differences in meaning, one must have a theory of meaning. After presenting four experiments to illustrate the pattern of results that must be interpreted, we reconsider the problem of meaning. In the General Discussion section, we suggest that the notion of distinctiveness of encoding should be used to supplement (or possibly to replace) that of depth of processing; the relationship of distinctiveness to meaning is also considered. In the discussion, the further point is made that an interpretation of effects in terms of distinctiveness requires that one consider the conditions of testing as well as those of study. Our present discussion expands on points made previously by Jacoby and Craik (1979).

The four experiments that follow explore the relations between decision difficulty and subsequent retention. The first study shows that correcting a word's spelling renders the word more memorable. Experiment 2 extends the idea of decision difficulty to a situation in which objects must be compared in greater or lesser detail before a specified decision can be made; again, more extensive initial processing was associated with higher retention levels. In Experiment 3, both the work required to compare two words and the words' preexperimental associative strength were manipulated in the initial decision phase. Both factors affected retention, but recognition and recall were affected in rather different ways. Finally, Experiment 4 also examined the interactions of decision difficulty, preexperimental accessibility, and type of retention test in determining performance. All factors affected the level of retention, in a complex but understandable way.

Experiment 1

The task employed in the first experiment required subjects to correct misspellings of the sort that are commonly encountered as typographical errors. A list of words was presented in which some of the words were misspelled. If an item was misspelled, the subject was to correct the spelling and write the correctly spelled version of the word; correctly spelled words were simply to be copied. A later retention test compared recognition of the correctly spelled version of previously correctly spelled and misspelled words.

Although spelling errors in this experiment were quite easy to correct, it was felt that the activity of correcting a spelling error might still benefit retention. Any retention advantage for previously misspelled words would be difficult to account for within the level of processing framework since correction of spelling and judging that a word is correctly spelled appear to require equally minimal processing of meaning. That is, differences in the processing of meaning do not provide a basis for predicting a retention advantage of words that were misspelled originally as compared to those that were correctly spelled.

The context in which a word appeared was also varied. Words for which spelling was to be checked and, if necessary, corrected appeared either alone or in the context of a related word; the context word was either similar in sound and appearance (taste paest) or similar in only sound (waist paest) to the correctly spelled version of the item. It was felt that the provision of context might further facilitate the correction of spelling errors and, consequently, reduce later recognition performance.

Method

Subjects. Twelve students at McMaster University were paid \$2.00/hour to participate. Testing was conducted in individual sessions.

Materials. Subjects were presented with a list of 40 words; the spelling of 20 of the 40 words was violated by interchanging two letters (e.g., reach raech). Subjects were instructed to copy correctly spelled words and to write down the correctly spelled version of each misspelled word. Twenty of the words (10 correctly spelled and 10 misspelled) were presented in the context of a similar item; for half the items of each class the context word was similar in both sound and appearance to the target word while for the other half it was similar only in sound. Four versions of the basic list were constructed so that across lists each of the four main conditions (spelled correctly vs misspelled × context vs no context) was represented by the same words. Each list of 40 items (20 correctly spelled and 20 misspelled words) was prepared as a deck of 3×5 -in. note cards; each word along with its context word (if any) appeared on a separate card. A different random order of the cards was prepared for each subject in the experiment.

A recognition test sheet contained the correctly spelled versions of the 40 words for which spelling had been checked plus 80 words that had not been previously presented. A second recognition sheet included the 20 items that had previously served as context words plus 60 words that had not been presented previously. On both recognition sheets, the order of items was randomly determined.

Procedure. The list of words was presented as a deck of note cards. Subjects were to work their way through this deck writing the correctly spelled version of each item on a sheet of paper. In addition, for each item encountered in the deck they were to give a rating of how difficult it was to check the spelling of that word. The rating scales provided for this purpose ranged from 1 to 7 with the subjects being instructed to use the number 1 to indicate that the spelling of an item was very easy to check. The task of working through the deck of cards was subject paced.

After the task of checking for spelling errors was completed, recognition tests were administered; subjects were not forewarned that these tests would be given. In the first recognition test, subjects were to circle items that had appeared in the deck of cards; further, they were to indicate for each item they circled whether that word had been correctly spelled or misspelled in the original list. In the second recognition test, subjects were to circle items that had previously served as context words. Both tests were subject paced.

The number of correct recognitions (hits) served as a measure of recognition for purposes of analyses; since items of the different types were intermixed in the recognition test there is no reason to believe that differences among conditions in number of hits were due to differences in response bias. The significance level for statistical tests was set at p < .05.

Results and Discussion

Only three subjects failed to correct all of the spelling errors in the original task. Of these three subjects, two accepted one misspelled word as being correctly spelled, while the third subject accepted three misspelled words as being correct.

The probability of recognizing items for which spelling had been corrected was substantially higher than that of recognizing items that had been correctly spelled as presented and merely copied (.73 vs .50), F(1, 22)=8.86, $MS_e=.07$. The probability of circling new items (false alarm) was quite low (.01). The higher recognition of incorrectly spelled words (which, we assume, required more extensive processing) is reminiscent of the effects recently reported by Slamecka and Graf (1978), and by Jacoby (1978).

The manipulation of context had no significant effect on later recognition. This lack of an effect occurred even though recognition of context words was fairly high; the probability of recognizing a word that had previously served as context was .42 while the probability of a false alarm was .03. The ineffectiveness of providing context may have been due to the spelling errors being so obvious that the provision of context did not substantially influence the ease of their correction. Alternatively, the relationship between decision difficulty and subsequent retention may be a discontinuous one so that increases in difficulty beyond some minimal level do not influence later retention (Jacoby, 1978, Experiment 2). Results of experiments to be reported later weigh against the latter alternative by showing a more continuous relationship between decision difficulty and retention.

Given recognition, the probability of correctly specifying that an item had been misspelled or correctly spelled in the original task was quite high (.87). Identification of the earlier spelling did not differ for correct spellings as compared to incorrect spellings, or as a function of manipulations of context.

EXPERIMENT 2

The second experiment was again designed to manipulate the degree of initial processing. In this case, however, items were chosen to be similar or dissimilar on a single dimension, namely, size of referent. Paivio (1975) has shown that pairs of words whose referent objects differ greatly in size lead to more rapid decisions about which is larger than do pairs with less difference; for example, it takes longer to make a size difference decision for tiger vs donkey than for frog vs kangaroo. The closer the two objects are in size, the more detailed processing must be in order to make the decision; consequently, pairs of similar-sized objects should yield more distinctive memory traces than do pairs of objects of very different sizes. In the example just given, subjects should have a higher probability of remembering the object paired with tiger than the object paired with frog.

Subjects rated pairs of nouns for the difference in size of referents. Nouns were selected from Paivio's (1975) norms, and pairs were constructed so that both members were from the same decile, or were separated by 2, 4, 6, or 8 deciles in the norms, with all possible deciles for a given separation equally represented. For example, with 0 separation, pairs were from the smallest decile (flea-ant), the 5th decile (kettle-football), and the largest decile (rhinoceros-garage), as well as all the others. Examples of items separated by 2, 4, 6, and 8 deciles are crumb-tomato, raisin-rabbit, roach-window, and bee-refrigerator; in the examples, the smaller item is always from the 1st decile.

Method

Subjects. Forty students were paid \$2.00/ hour to participate. Testing was conducted in individual sessions of 30-minute duration.

Materials. Two lists of noun pairs were constructed, with 80 pairs in one list, and 72 in the other. The nouns were selected from Paivio's (1975) norms, which provide for each noun the mean rated size of its referent object. Based on the norms, the nouns were sorted into deciles, with about 16 nouns in each. The 80-pair list contained 40 pairs in which both members were selected from the same decile, with 4 pairs from each decile; another 32 pairs contained one member from decile *n* and another from decile n+2, with *n* ranging from 1 to 8, and 4 pairs of each possible value; similarly, 8 pairs contained one member from decile *n* and one from n+8, with n equaling 1 or 2, and 4 pairs of each value. The 72-pair list was also constructed with 4 pairs of each range, with 32 members separated by 2 deciles, 24 by 4, and 16 by 6.

Each pair was printed on a separate index card, yielding two decks of study material. Additionally, a test list was constructed for each study deck, with one randomly selected member of each pair appearing as a cue.

Procedure. Each subject received a deck of cards, and rated the difference in size of the named objects, using a scale from 1 (no difference) to 10 (vast difference). After the rating task was completed, the cued recall test was administered. Presentation and test were self-paced, with 20 subjects receiving each list.

Results and Discussion

The mean ratings produced the predictable results, with rated difference increasing with scaled difference. Items separated by 0, 2, and 8 deciles produced ratings of 2.00, 4.25, and 8.02, with ratings of 4.29, 5.85, and 7.03 for the items separated by 2, 4, or 6 deciles, respectively. More importantly, mean proportion recalled declined regularly with the size difference, with means of .27, .15, and .10 for the pairs separated by 0, 2, or 8 deciles, F(2, 38) = 34.3, $MS_e = .0044$, and means of .17, .11, and .09 for pairs separated by 2, 4, or 6 deciles, respectively, F(2.38) = 11.3, $MS_e = .0033$. Quite clearly, more detailed but equally meaningful initial processing was associated with higher levels of subsequent recall.

EXPERIMENT 3

In the third experiment, we again demonstrate that a more difficult decision enhances later retention performance. More obviously than in the earlier experiments, perhaps, the task employed in Experiment 3 was one that clearly required semantic processing in all conditions. One point to be made by the third experiment is that differences in retention performance can be produced by a manipulation of decision difficulty even when all of the conditions that are being compared require the processing of meaning. Other points to be made concern the role of retrieval processes; the effectiveness of retrieval was manipulated by varying the form of the retention test.

Subjects were required to judge which of two alternatives held the stronger associative relationship with a focus word; the difficulty of the decision was varied by manipulating the degree of association between each of the alternatives and the focus word. For example, when the focus word was "water," the decision would be a difficult one if the alternatives were "lake" and "thirst" (both high associates of "water"). In contrast, the decision would be an easy one if the alternatives were "lake" and "chair" (a high associate and a word that is unrelated to "water"). Generally, the more difficult decision was expected to result in higher performance on a test of incidental retention.

Method

Subjects. Thirty-six students were paid \$2.00/hour to participate. Testing was con-

ducted in individual sessions.

Task and procedure. Subjects judged relationships among words presented on index cards. Each card had one word (the focus word) printed on one side, and two words printed on the reverse side. The subject's task was to study the focus word, and then turn the card over and pick which word from the two alternatives was more highly related to the focus word. The two alternatives were both highly related to the focus word (High-High), or one high associate and one low associate of the focus word (High-Low), or one high associate and one word that was unrelated to the focus word (High-Unrelated). Similarly, other conditions were Low-Low, Low-Unrelated, and Unrelated-Unrelated. A deck contained 60 index cards; 10 cards represented each of the 6 conditions defined by difference in degree of relationship between the two alternatives and the focus word.

Subjects were told that the purpose of the experiment was to determine what factors influence the speed with which relationships among words are judged. They were then instructed to work their way through the deck of cards, as rapidly as possible, saying aloud the alternative on each card that was more strongly related to the focus word. After they had gone through the deck of cards, subjects were given an unexpected test of retention. Subjects in one condition were provided with the focus words as cues for recall of the alternatives that they had chosen as being most strongly related to the focus words. Subjects in a second condition were given a test for recognition of the alternatives that they had chosen as being most strongly related to the focus words. This procedure allowed the comparison of recognition and recall of the same items-alternatives that had been selected during the first part of the experiment. Presentation of the study list and all forms of test were subject paced.

Materials. A set of 60 words was chosen from the Connecticut Free Associational Norms (Bousfield, Cohen, Whitmarsh, &

Kincard, Note 1) to serve as focus words. Two words from the five most frequent associates to each focus word were selected as high associates; two words that were among the associations that were given by only one subject were selected as low associates. This set of focus words, high associates, and low associates was arranged to produce 10 instances each of the 6 judgment conditions described earlier: High-High, High-Low, High-Unrelated, Low-Low, Low-Unrelated, and Unrelated–Unrelated. An unrelated item within these judgment conditions was actually a high associate of a focus word in the list other than that with which it was presented; in the Unrelated-Unrelated case, the two "unrelated" words were never high associates of the same focus word. Further, an "unrelated" word was always separated by at least three cards from the focus word to which it was a high associate. With this restriction, the assignment of sets of items to list positions was random. Six replications of the list were produced by rotating focus words through conditions so that, across replications, each focus word represented each of the six judgment conditions. To the extent possible, alternatives were also maintained across replications. For example, the High alternatives in the High-Low and in the High-Unrelated conditions comprised one-half of the words used as alternatives in the High-High condition.

For the cued recall test, the focus words were typed in a random order with a blank next to each focus word; subjects were to fill each blank with the alternative that they had chosen as being most strongly related to the focus word during presentation of the list. For the recognition test, the experimenter composed a deck of note cards containing each alternative the subject had chosen during the judgment task. This deck of 60 note cards was then randomly intermixed with 120 note cards that contained "new" items. The new items included 60 words that were high associates of the focus words; these items were selected from those words that were among the five most frequent associates of a focus word, but that had not been used in constructing the original study list. The remaining 60 new items were selected to be unrelated to any focus word or alternative presented in the judgment task. Subjects worked through the deck of cards comprising the recognition test, writing down those words that they thought had occurred in the original judgment task.

Analyses. In general, the analyses included form of the retention test as one factor and either decision difficulty or strength of association between the alternatives and the focus word as a second factor. The effects of decision difficulty were assessed by means of comparisons that held the associative strength of the chosen alternative constant while varying the associative strength of the nonchosen alternative (i.e., High-High vs High-Low vs High-Unrelated, and Low-Low vs Low-Unrelated). The effects of strength of association were assessed by holding the similarity between alternatives constant while the strength of their association with the focus word varied (High-High vs Low-Low vs Unrelated–Unrelated). The significance level for all tests as set at p < .05.

The number of correct recognitions served as a measure of recognition performance for the primary analyses. Other analyses compared the number of false recognitions for the various conditions; this comparison of false recognitions was possible only with "new" items that were associatively related to a focus word.

Results and Discussion

The main results are shown in Fig. 1. Inspection of those results reveals that both decision difficulty and the strength of the normative association between the focus word and the alternatives had strong effects on retention performance. Further, each of those variables interacts with the form of the retention test. Both the effect of decision difficulty and that of prior associative strength are more pronounced in cued recall than in recognition. Finally, it should be noted that cued recall scores are higher than recognition scores for the highly associated materials, but that the superiority of cued recall drops for low associates, and reverses for unrelated words.

Statistical evidence of the effects of decision difficulty comes from comparisons that equate the associative strength of the chosen alternative while varying the associative strength of the not-chosen alternative. When the pair of alternatives included a high associate of the focus word (HH, HL, and HU), cued recall and recognition increased as the similarity in associative strength of the two alternatives increased, F(2, 68)=9.00, $MS_e=.02$. Further, cued recall and recognition were greater in the Low-Low condition than in the Low-



FIG. 1. Recognition and cued recall as a function of associative strength of test alternatives.

Unrelated condition, F(1, 17) = 30.1, $MS_e = .01$. Although the effect of decision difficulty was numerically larger in cued recall than in recognition, the interaction of form of test with decision difficulty did not attain significance in either of the two above sets of comparisons.

The effects of strength of association between the focus word and alternatives can be seen by comparing performance in the High– High, Low–Low, and Unrelated–Unrelated conditions, conditions in which decision difficulty is roughly equated. Both cued recall and recognition of the chosen alternative decline as the strength of the association between the focus word and the alternatives is decreased, F(2, 68)=74.5, $MS_e=.02$. However, as shown by the significant interaction with form of test, F(2, 68)=20.5, $MS_e=.02$, the magnitude of the decline across level of associative strength was much more pronounced in cued recall than in recognition.

Analyses of errors provides a means of assessing the precision of recognition and recall performance. For example, it might be argued that presenting two high associates of a focus word enhances later recognition by making it more likely that the subject will remember the general concept represented by the focus word; if this is so, the subject in the High-High condition might be more likely to mistake new high associates of focus words as being old.

Analyses of the recognition errors did reveal that subjects were more likely to mistake new high associates of a focus word as being old (.04) than they were to mistake new items that were unrelated to focus words as being old (.01). However, further analyses of the new high associates did not reveal any significant effect of decision difficulty on the probability of a false recognition. Indeed, numerical differences were in the opposite direction to what would be expected if the more difficult decision led to a more general encoding; for example, the probability of a false recognition of a new high associate was only .01 in the High–High condition while it was .06 in the High–Unrelated condition.

Other analyses examined the probability of an intrusion error in cued recall. Overall, the probability of an extralist intrusion was .10; the probability of an extralist intrusion did not differ as a function of either decision difficulty or strength of association between the focus word and alternatives. The probability of an intralist intrusion, in contrast, was higher when at least one of the alternatives was unrelated to the focus word (.09) than when both of the alternatives were related to the focus word (.04). This higher rate of intralist intrusions for unrelated items probably reflects the fact that unrelated items were actually high associates of focus words other than those with which they were presented; recall of an "unrelated" item with the cue to which it was a high associate was counted as an intrusion error.

The effects of decision difficulty obtained in the present experiment agree with the results of other experiments (e.g., Begg, 1978; Klein & Saltz, 1976; Packman & Battig, 1978) by showing differences in retention among conditions that require the processing of meaning. One way to account for these differences among conditions is to suggest that although all conditions processed meaning, the conditions differed in their "elaboration" of the presented items (Craik & Tulving, 1975). A second account that will be expanded upon in the General Discussion section adopts a view of meaning that differs from that held by Craik and Lockhart (1972). This alternative view treats meaning as arising from the contrasts or distinctions conveyed by the use of a word in the context of a particular task. The problem of determining the meaning of an item is seen as being analogous to the problem of describing an object. As pointed out by several authors (e.g., Brown, 1958; Garner, 1974; Olson, 1970), what an object is called or how it is described depends on the other objects from which it is to be discriminated. For example, a chair is a chair but it is equally a piece of furniture, a thing, a wooden object, and any number of other descriptions, depending on what the chair is to be distinguished from. Similarly, the meaning of a word depends upon distinctions that are to be conveyed by that word in that context. By this view of meaning, the more difficult decisions in the present experiment required further distinctions to be drawn, thereby resulting in more precise "descriptions" of the presented words.

Since the focus words are given as cues for recall of the target words, the normative strength of the association between the focus word and the target is particularly important for cued recall. In contrast, recognition was less influenced by variations in associative strength. However, the finding that recognition is somewhat affected by the strength of association between the focus word and chosen alternative, as well as by decision difficulty, suggests that in some instances retrieval processes are "expanded" to make use of the focus-target interactions that took place at encoding. The crossover between recall and recognition levels emphasizes that retention level depends upon both study encoding and the effectiveness of the retrieval cue to reconstitute the encoded information at the time of test. This last point is taken up again in the fourth and final experiment.

EXPERIMENT 4

The fourth experiment also varied the difficulty of the initial judgment task and the type of retrieval information provided at test. The initial judgment task consisted of answering a category question, such as "Is X a type of animal?" The test items and category labels were selected from the Battig and Montague (1969) norms such that the target words differed in the frequency with which they were produced as members of the category. Note, however, that the words were equated for frequency of occurrence in the language. After the initial task was completed,

half the subjects were provided with the category names and asked to recall the target words, while the other half were given a recognition test of their memory for the target items.

The expectations regarding relative levels of retention are not immediately obvious. Since it was expected that less common exemplars would require more extensive processing before a decision regarding category membership could be made, such exemplars should give rise to more distinctive traces. On the other hand, it is known from the norms that more common exemplars are more readily retrieved when the category name is provided as a retrieval cue. It is argued that these opposing effects may differentially affect recall and recognition. Speculatively, the process of recall relies heavily on selfgenerated reconstructive operations whereas the process of recognition involves reconstruction to a lesser extent. If this is so, it may be argued that when a subject attempts to recall previously presented category exemplars, the reconstructive operations evoked are similar to those involved in the production of category norms, and that common exemplars are thus more readily recalled than are less common exemplars, despite the less distinctive traces of common items. In recognition, however, the subject must choose presented items from a list containing other exemplars of the same category; therefore, items with more distinctive memory traces should yield better discrimination than items with less distinctive traces. By this argument, recognition performance should be inversely related to the normative probability of giving the word as a response to the category name (Schnur, 1977).

The expected interaction, with recall positively related to dominance, but recognition negatively related, has been obtained previously by Rabinowitz, Mandler, and Patterson (1977b). However, in their experiments, the interaction was obtained only when frequency of occurrence was uncontrolled. When they controlled for frequency, no difference in free recall occurred across taxonomic frequency; recognition tests were not given. Accordingly, the matter needs further clarification.

Method

Subjects. Thirty-two students performed the experiment for a payment of \$3.00. Each subject was tested individually.

Materials and procedure. The subject was informed that the experiment was concerned with the ability to decide rapidly whether or not a word belongs in a specified category. On each trial the experimenter first spoke the category name, then 2 seconds later the target word was exposed for 300 milliseconds on a tachistoscope screen. The subject responded "yes" or "no" by pressing one of two telegraph keys in front of him; reaction times were recorded. There were 72 trials in all, of which 48 required a positive response and 24 required a negative response. The 48 positive words were made up of 3 examples from each of 16 categories. The 24 negative cases used words drawn from 24 new categories, each paired with an inappropriate category name. The negative cases were included merely to make the initial task plausible, and are not considered further. After completing the 72 decision trials, the subject was asked to return the following day to complete the experiment. In this second session half of the subjects were given a cued recall test for the 48 positive words (the 16 category names served as cues), and the other half of the subjects were given a recognition test for the 48 positive words, mixed randomly with 144 nonpresented words from the same 16 categories. Further details of the retention tests are given below; the recall and recognition tests were administered on the following day to eliminate ceiling effects.

The 48 positive trials involved target words drawn from 16 categories in the Battig and Montague (1969) norms. Three exemplars were drawn from each category; one from the top, middle, and bottom thirds of each list. Thus, of the 48 words, 16 were frequently generated, 16 were intermediate, and 16 were infrequently generated members of the categories used. However, despite these differences in frequency of generation, the three exemplars chosen from each category were equated for frequency in the language (Thorndike & Lorge, 1944).

On the following day, 16 subjects were provided with the 16 category names as cues and asked to write down as many as possible of the 3 words in each category. The remaining 16 subjects were given a recognition test for the 48 target words. The targets were typed on a sheet, mixed randomly with 144 distractor words drawn from the same 16 categories. Nine distractor words were drawn from each of the top, middle, and bottom thirds. In recall, two orders of presentation of the category name cues were used, one order for eight subjects and the other order for the remaining eight. Similarly, in recognition, two random orders of the targets plus distractors were used. In the recognition test, subjects were asked to check exactly 48 items as old. In both recognition and recall, testing was self-paced.

Results and Discussion

Initial decision times may be taken as an approximate measure of decision difficulty. Reaction times to the High, Medium, and Low category exemplars were 813, 871, and 914 milliseconds, respectively. This effect was statistically reliable, F(2, 60) = 18.37, $MS_e = 4463$. Thus it may be concluded that decision difficulty increased from High to Medium to Low category exemplars.

Figure 2 shows that recognition rates increased from High to Medium to Low category exemplars, whereas cued recall was highest for the most common category members. The interaction between recall vs recognition and position in the category was reli-



FIG. 2. Recognition and cued recall as a function of taxonomic frequency.

able, F(2, 60) = 15.1, $MS_e = .02$. Thus, decisions about category membership had rather different effects on recognition and recall in this situation. We argue that the greater difficulty of initial decision associated with uncommon category members yielded more distinctive traces, and thus higher levels of recognition performance. In recall, however, the ease of generation of the target words from the category name played a more dominant role; High category exemplars produced the highest recall levels despite the assumption that their traces were less distinctive. In summary, the argument is that recognition levels reflect trace distinctiveness, but that distinctiveness can be overcome by another factor-ease of reconstruction-in recall.

GENERAL DISCUSSION

The results of the four experiments just reported agree with the conclusion of several recent investigators that, within the domain of semantic processing, decision or response difficulty is positively related to subsequent memorability (Auble & Franks, 1978; Epstein, Phillips, & Johnson, 1975; Jacoby, 1978; Kolers, 1973; Schnur, 1977; Slamecka & Graf, 1978). What further conclusions can be drawn from the present experiments and how do these conclusions address current notions of meaning and of memory?

At the empirical level, the experiments have demonstrated that other factors besides

the difficulty of initial processing must be taken into account. Whereas Experiment 1 showed that correcting a spelling error rendered a word more memorable than the act of simply copying the word (the first task presumably involved more cognitive effort than the second) and Experiment 2 showed that retention was related systematically to the difficulty of an initial comparison, Experiments 3 and 4 brought other factors into play. In these latter studies, retention was assessed both by cued recall and by recognition; also, the preexperimental strength of relation between the target item and the comparison word or category was varied. In Experiment 3, difficulty of decision and preexperimental relatedness were varied orthogonally; in Experiment 4, the two variables were placed in opposition. Experiment 3 showed that preexperimental strength correlated positively with both cued recall and recognition, but that the effect was much stronger in the case of recall. Experiment 4 showed that when decision difficulty and preexperimental strength were pitted against each other, recognition level was affected primarily by decision difficulty, whereas preexperimental strength was again an important determinant of recall. However, preexperimental strength had some effects on recognition performance (Experiment 3), and decision difficulty clearly affected recall in Experiments 2 and 3 (and perhaps in Experiment 4 also, since slightly more Low exemplars were recalled than Medium exemplars).

Thus the four experiments yield a highly consistent pattern of results. To understand the effects observed in the situations used in the studies, the factors of decision difficulty, type of test, and the preexperimental strength between some aspect of the context and the target item must all be borne in mind. Decision difficulty is associated with higher retention levels for both recognition and recall. Preexperimental strength also affects both indices of retention, but has a greater effect on recall.

Later in this section we develop the view that difficult initial processing implies more extensive or elaborate analysis, and that this more extensive analysis is reflected in a richer, more distinctive memory record of the event. The distinctive record, in turn, is highly discriminable from other memory traces and is retrieved with relative ease. Successful retrieval apparently requires that some critical number of attributes or operations activated during the retention test match those anticipated at the time of initial encoding (Eysenck, 1979; Flexser & Tulving, 1978; Kolers, 1973). Since only a portion of study context is perceptually available in recall, recall must be accomplished by some regenerative means such as generation-recognition (Bahrick, 1970), reconstruction (Lockhart, Craik, & Jacoby, 1976), or redintegration (Begg, 1972; Horowitz & Prytulak, 1969). In contrast, recognition may rely less on reconstructive operations because a sufficient match between past and present operations can be "driven" by the represented stimulus itself. That is, recognition may often be relatively "context free." The suggestion, then, is that successful recall depends heavily on the potential for reconstruction of the event, whereas recognition is more dependent on the distinctiveness of the memory record.

This account of the differences in emphasis between recall and recognition provides a way to interpret the observations that whereas both recognition and recall benefit from difficult initial processing, recall is more affected by strong preexperimental relations between cue and target (thereby facilitating reconstructive operations) and recognition is more affected by decision difficulty. However, it is clear that recognition is not generally a context-free process when recognition of specific episodes is being tested. In many cases reconstructive processes will come into play in recognition as well as in recall; that is, retrieval processes in recognition will be "expanded" to achieve a fuller redintegration of the initial context in which the event occurred. In these cases strong preexperimental relations between the item and some aspects of the context will serve to enhance the level of recognition (Experiment 3). The distinction between "context-free" and "expanded" processing in recognition is similar to Mandler's distinction between presentation and conceptual codes in recognition (Mandler, 1972; Rabinowitz, Mandler, & Barsalou, 1977a). By this account, if the relation between cue and target is made particularly strong, either normatively (Experiment 3) or episodically (Tulving & Thomson, 1973), the finding of higher cued recall than recognition is not at all surprising.

Finally, how exactly does greater difficulty of the initial decision enhance retention? As outlined previously, we endorse the suggestion that perception may be viewed as the process of describing the stimulus (Norman & Bobrow, 1979; Rock, 1975) and that the memory trace may be regarded as the record of this description. The more complex and difficult the initial task, the richer, more elaborate, and more precise the resulting description. In turn, precise descriptive records are likely to be distinctive and potentially retrievable, provided effective cues are given to guide retrieval processing. In Experiments 2, 3, and 4, the meaning of words to be remembered was clearly involved in all conditions. Consequently it would be difficult to attribute the observed differences in retention to differences in "levels" of processing (Craik & Lockhart, 1972). Whereas it seems clear that the analysis of meaning is not automatic and does not occur in an all-or-none fashionthat is, semantic analysis can vary from cursorv to detailed (Johnson-Laird et al., 1978)--the notion that such variations constitute qualitatively different stages or levels of analysis has little intuitive appeal. Within the domain of semantic processing, at least, it may be more fruitful to relate retention to the richness of the descriptive record and thus to the distinctiveness of that record.

However, distinctiveness is a relative, not

an absolute, term; an object or description is distinctive relative to some particular background. In the present case, also, the distinctiveness of a memory record will always be relative to a given context, and the same context must be reinstated at retrieval if the encoded distinctiveness of the memory trace is to be optimally utilized. Accordingly, distinctiveness cannot be discussed without reference to conditions of study and conditions of testing. It is this relativity of the memory record to its context that distinguishes our use of the term "distinctiveness" from "elaboration" (Craik & Tulving, 1975). Elaboration often refers to the addition of further information, so that the trace becomes richer and more detailed. In using "distinctiveness," however, we mean to emphasize the contrastive value of information in the trace.

The original levels of processing view was based implicitly on the assumption that each word has a single basic meaning, or at most a few meanings. Therefore, it was reasonable to consider meaning as a level of processing, or something which is either encoded or not encoded. The effect of different orienting tasks and context was seen as influencing the likelihood that the meaning of a stimulus word is registered during processing. This basic fixedentity view of meaning corresponds well with common sense and behavioristic conceptions of language. For example, Bloomfield (1933) argued that each word has a single core meaning, although additional marginal meanings may exist. Similarly, Katz and Postal (1964) adopted a fixed-entity view of meaning, with syntactic markers, semantic markers, and distinguishers serving as three different levels of generality of the fixed entities.

Despite the plausibility of the fixed-entity view, there are situations in which it is more reasonable to consider meaning as defined (or at least as strongly influenced) by context, rather than as selected by the context. As pointed out earlier in the present article, the descriptive label chosen to describe a

chair depends critically on what the chair is to'be distinguished from. Similarly, the meaning of a word in a given context depends heavily on the distinctions to be conveyed by the word in that context. That is, the meaning of a word is a range of potential, rather than a set of encodings. Another example of the dependence of meaning on the distinctions to be conveyed can be seen in considering synonymy. On the one hand, there are few true synonyms in English, if by synonymy we mean that two words can be used interchangeably in all contexts with no consequence for interpretation (e.g., Herrman, 1978). On the other hand, it is possible to find contexts in which many different words could be used without changing the essential sense of the statement. For example, in the context "The actor was so short that he had to stand on a _____ to kiss his leading lady," such diverse words as "box," "stool," and "book" can be used interchangeably. The point is that meaning is not a static characteristic of individual words, but is rather a variable set of interpretations depending on the distinctions the word is intended to convey. Wittgenstein's advice, to ask not what a word means but rather how it is used, has recently also been accepted by Bransford, Franks, Morris, and Stein (1979), Anderson and Ortony (1975), and others.

If we are to consider meaning in terms of the distinctions a word conveys, we cannot treat meaning as a single level of processing. Processing or interpreting a word in any task or context requires the drawing of some distinctions. Consequently, in accounting for different levels of retention, we are led to focus on differences in the number of distinctions, the nature of distinctions, and the utility of distinctions required in initial processing tasks. As mentioned earlier, a parallel situation has occurred in perception, in which several theorists have come to view perception as the process of describing the stimulus (e.g., Norman & Bobrow, 1979; Rock, 1975). Since contextual factors such as the alternatives from which a stimulus is to be discriminated influence the description, they necessarily influence perception. If the memory trace is the record of perceptual analyses, it too can be treated functionally as a description or set of contrasts. The resultant view of a memory trace is similar to a multicomponent (Bower, 1967) or attribute (Underwood, 1969) theory, differing primarily in the emphasis placed on the extent to which the description is necessarily relative to a given context, since context specifies the appropriate contrasts.

The distinctiveness of a memory trace, rather than its depth, can be used to account for a variety of retention differences. Retention will be limited by the relation between distinctions drawn during study and those required at the time of test. For example, if "lady" occurred in a study list, we might expect many false recognitions of "woman" during test (e.g., Underwood, 1965). However, if the initial task required drawing distinctions about deportment as well as age and sex, fewer errors should result. In short, the alternatives with which an item will be confused depend on how the word was initially encountered and encoded. In general, the more complete and precise the initial description is, the smaller the set of confusable alternatives will be.

Let us stress again that distinctiveness is context sensitive; a description that is highly distinctive for one set of alternatives may not be distinctive for another set. If "beer" is initially encoded as an alcoholic beverage, it may be poorly distinguished from "wine" or "vodka," but if "beer" is encoded as a good thirst quencher after baseball, it may be poorly distinguished from orange juice or Coca-Cola. Consequently, it is impossible to specify the distinctiveness of the description of an event without specifying the alternatives in question. Any major change in the set of alternatives may make a previously distinctive description useless. The relativity of similarity and distinctiveness in perceptual and conceptual contexts has recently been emphasized by Tversky (1977). In memory theory also, there is growing realization that the effectiveness of a particular retrieval context is not absolute, but depends on the way in which the event was encoded (Begg, 1978; Bransford, Franks, Morris, & Stein, 1979; Tulving & Thomson, 1973). Such views necessarily focus on interactions between the manner in which initial encoding was carried out and the demands of the particular situation in which memory is assessed.

In summary, the present article has presented further evidence for the importance of initial decision difficulty for later retention of the event. We have argued that greater initial difficulty does not involve "deeper levels" of encoding, but rather that greater difficulty necessitates more extensive processing of the event with the concomitant formation of a more precise perceptual "description" and a more distinctive memory record. Distinctiveness enhances memory in the same way that contrast enhances perception-by increasing the informational overlap between the item and its perceptual specification (or the retrieval information provided) and by decreasing the informational overlap between the item and others in the visual field (or in the memory system) (Eysenck, 1979; Flexser & Tulving, 1978; Tversky, 1977). In addition, we have argued that effective retrieval conditions serve to reinstate the background against which distinctiveness is accomplished, and that factors facilitating reconstructive processing are particularly important in the case of recall.

REFERENCES

- ANDERSON, R. C., & ORTONY, A. On putting apples into bottles—A problem of polysemy. *Cognitive Psychology*, 1975, 7, 167–180.
- AUBEL, P. M., & FRANKS, J. J. The effects of effort toward comprehension on recall. *Memory and Cognition*, 1978, 6, 20–25.
- BADDELEY, A. D. The trouble with levels: A reexamination of Craik and Lockhart's framework for memory

research. Psychological Review, 1978, 85, 139-152.

- BAHRICK, H. P. A two-phase model for prompted recall. *Psychological Review*, 1970, **77**, 215–222.
- BATTIG, W. F., & MONTAGUE, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal* of Experimental Psychology, 1969, **80**, (3, Part 2).
- BEGG, I. Recall of meaningful phrases. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 431–439.
- BEGG, I. Similarity and contrast in memory for relations. Memory and Cognition, 1978, 6, 509–517.
- BLOOMFIELD, L. Language. New York: Holt, Rinehart & Winston, 1933.
- BOWER, G. H. A multicomponent view of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The* psychology of learning and motivation: Advances in research and theory. New York: Academic Press, 1967. Vol. 1.
- BRANSFORD, J. D., FRANKS, J. J., MORRIS, C. D., & STEIN, B. S. Some general constraints on learning and memory research. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1979.
- BROWN, R. W. How shall a thing be called ? Psychological Review, 1958, 65, 14–21.
- 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.
- EPSTEIN, M. L., PHILLIPS, W. D., & JOHNSON, S. J. Recall of related and unrelated word pairs as a function of processing level. *Journal of Experimental Psycholo*gy: Human Learning and Memory, 1975, 1, 149–152.
- EYSENCK, M. W. Levels of processing: A critique. British Journal of Psychology, 1978, 69, 157–169.
- EYSENCK, M. W. Depth, elaboration and distinctiveness. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1979.
- FLEXSER, A. J., & TULVING, E. Retrieval independence in recognition and recall. *Psychological Review*, 1978, 85, 153-171.
- FRASE, L. T., & KAMMANN, R. Effects of search criterion upon unanticipated free recall of categorically related words. *Memory and Cognition*, 1974, 2, 181–184.
- GARNER, W. R. The processing of information and structure. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.
- HERRMANN, D. J. An old problem for the new psychomantics: Synonymity. *Psychological Bulletin*,

1978, 85, 490-512.

HOROWITZ, C. M., & PRYTULAK, L. S. Redintegrative memory. *Psychological Review*, 1969, **76**, 519–531.

- JACOBY, L. L. On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, 1978, 17, 649-667.
- 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* and human memory. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1979.
- JOHNSON-LAIRD, P. N., & BETHELL-FOX, C. E. Memory for questions and amounts of processing. *Memory* and Cognition, 1978, 6, 496-501.
- JOHNSON-LAIRD, P. N., GIBBS, G., & DE MOWBRAY, J. Meaning, amount of processing, and memory for words. *Memory and Cognition*, 1978, 6, 372–375.
- KAHNEMAN, D. Attention and effort. Englewood Cliffs, N.J.: Prentice-Hall, 1973.
- KATZ, J. J., & POSTAL, P. An integrated theory of linguistic descriptions. Cambridge, Mass.: M.I.T. Press, 1964.
- 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.
- KOLERS, P. A. Remembering operations. *Memory and Cognition*, 1973, **1**, 347–355.
- KOLERS, P. A. Memorial consequences of automatized encoding. Journal of Experimental Psychology: Human Learning and Memory, 1975, 1, 689–701.
- LOCKHART, R. S., CRAIK, F. I. M., & JACOBY, L. L. Depth of processing, recognition and recall: Some aspects of general memory system. In J. Brown (Ed.), *Recall* and recognition. London: Wiley, 1976.
- MANDLER, G. Organization and recognition. In E. Tulving & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.
- MANDLER, G., & WORDEN, P. E. Semantic processing without permanent storage. *Journal of Experimental Psychology*, 1973, **100**, 277–283.
- NELSON, T. O. Repetition and depth of processing. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 151-172.
- NORMAN, D. A., & BOBROW, D. G. Descriptions. An intermediate stage in memory retrieval. *Cognitive Psychology*, 1979, **11**, 107–123.
- OLSON, D. R. Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review*, 1970, 77, 257–273.
- PACKMAN, J. L., & BATTIG, W. F. Effects of different kinds of semantic processing on memory for words. *Memory and Cognition*, 1978, 6, 502–508.
- PAIVIO, A. Perceptual comparisons through the mind's eye. *Memory and Cognition*, 1975, **3**, 635–647.

- RABINOWITZ, J. C., MANDLER, G., & BARSALOU, L. W. Recognition failure: Another case of retrieval failure. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 639–663. (a)
- RABINOWITZ, J. C., MANDLER, G., & PATTERSON, K. E. Determinants of recognition and recall: Accessibility and generation. *Journal of Experimental Psychology: General*, 1977, **106**, 302–329. (b)
- ROCK, I. An introduction to perception. New York: Macmillan Co., 1975.
- SCHNUR, P. Testing the encoding elaboration hypothesis: The effects of exampler ranking on recognition and recall. *Memory and Cognition*, 1977, **5**, 666–672.
- SLAMECKA, N. J., & GRAF, P. The generation effect: Delineation of a phenomenon. Journal of Experimental Psychology: Human Learning and Memory, 1978, 4, 592-604.
- THORNDIKE, E. L., & LORGE, I. *The teacher's word book* of 30,000 words. New York: Columbia University, 1944.

- TULVING, E., & THOMSON, D. M. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 1973, **80**, 352–372.
- TVERSKY, A. Features of similarity. *Psychological Review*, 1977, **84**, 327–352.
- UNDERWOOD, B. J. False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, 1965, **70**, 122–129.
- UNDERWOOD, B. J. Attributes of memory. *Psychological Review*, 1969, **76**, 559–573.

REFERENCE NOTE

 BOUSFIELD, W. A., COHEN, B. H., WHITMARSH, G. A., & KINCARD, W. D. *The Connecticut Free Associational Norms* (Tech. Rep. No. 35). 1961.

(Received January 17, 1979)