Age-of-acquisition effects in lexical and episodic memory tasks

K. J. GILHOOLY and M. L. GILHOOLY
Aberdeen University, Aberdeen, Scotland, AB9 2UB, United Kingdom

This paper reports four experiments on the effects of word age of acquisition in verbal tasks. In all cases, multiple-regression analysis was used to assess the relative effects of age as opposed to other potentially relevant word attributes. Experiments 1 and 2 concerned lexical memory tasks. In Experiment 1, picture naming speeds were found to be mainly determined by picture codability and name age of acquisition. In Experiment 2, it was found that when subjects produced words in response to bigram cues, early acquired target words were more likely to be produced than later acquired words, even when frequency and other word attributes were taken into account. The remaining two experiments dealt with the episodic memory tasks of free recall and recognition. No age effects were found in these tasks. It was concluded that early age of acquisition facilitates retrieval from lexical memory but has no significant effect in episodic memory tasks.

This paper concerns the effects of word age of acquisition in lexical and episodic memory tasks. The first two experiments reported here involve retrieval from lexical memory in response to pictorial and letter cues, respectively. The remaining two experiments deal with the episodic memory tasks of free recall and recognition memory. Before detailing these experiments, we briefly review previous work on age-of-acquisition effects.

In one of the first studies in this area, Carroll and White (1973) found that the rated age of acquisition of picture names was a major determinant of picture naming latency. (It may be noted that the age ratings were highly reliable and correlated well with objective indices of age of acquisition.) In Carroll and White's study, early acquired names were generally produced faster than later acquired names. Further, this effect was not due to the correlation between age and frequency of usage, since age still had a significant correlation with picture naming speed when the effects of name frequency were partialled from age, while the correlation of frequency with speed dwindled to nonsignificance when age was partialled from frequency. This result was of potentially far reaching importance, since it suggested that the many word frequency effects reported in memory, learning, and perception research might be entirely due to the underlying, and hitherto largely unnoticed, variable of age of acquisition. However, subsequent studies (Lachman, 1973; Lachman, Schaffer,

& Hennrikus, 1974) found that both age and frequency can independently affect picture naming latency.

The experiments of Lachman and his co-workers are notable for the control exercised over the stimulus "codability" variable, a variable that Carroll and White (1973) did not include in their analyses. The codability of a picture for a given language community is the degree of consensus on the name to be given to the picture in that community. Highly codable pictures were found by Lachman to be named faster than pictures low in codability. Lachman et al. (1974) found that high codability was associated with low age-of-acquisition names, and this result was confirmed by Butterfield and Butterfield (1977). Lachman's (1973) regression analyses of individual data indicated that age, frequency, and codability are important independent factors in picture naming.

Taken together with Loftus and Suppes' (1972) findings of age effects in category-instance naming, the above results suggest that age of acquisition might be an important variable in a wide range of word retrieval tasks. Following up such an idea, Stratton, Jacobus, and Brinley (1975) examined age effects in anagram solving, a task that may be interpreted as involving retrieval to implicitly generated letter cues. Stratton et al. reported that rated age was a better predictor of anagram difficulty than the rated familiarity, objective frequency, imagery, or meaningfulness of the solution word. However, Gilhooly and Johnson (1978) found that, while rated age had a significant simple correlation with anagram solution probability, the effect dwindled when measures of the letter-order predictability of the solution word (Mendelsohn, 1976) were partialled out. The Gilhooly and Johnson study indicated a tendency for words rated as acquired early to have relatively predictable letter structures, and similar results have

been found for a larger word sample (Gilhooly, 1978). Just as high- and low-frequency words appear to differ in letter structure (Kinsbourne & Evens, 1970; Landauer & Streeter, 1973), so also do high and low age-of-acquisition words.

The correlations between age and letter-structure factors raise the possibility that some other apparent age effects, outside the anagram area, could also be due to letter-order factors. In particular, picture naming speeds have been measured by timers stopped by voice keys, and time to organize articulation of the name could be correlated with letter-order predictability, so producing an apparent age effect. Experiment 1 was designed to examine the role of age of acquisition in picture naming with letter-order factors taken into account.

If anagram solving involves retrieval to implicitly generated letter cues, then Gilhooly and Johnson's (1978) results might be taken to mean that age of acquisition does not affect retrieval to letter cues. Alternatively, it may be that word retrieval to implicit letter cues was not a major factor in Gilhooly and Johnson's task compared to, say, letter manipulation processes. Interpretation would be clearer if there was some direct evidence on age-of-acquisition effects on retrieval to letter cues. Experiment 2 was devised to provide information on this point by using a word completion task in which the subject was given a starting bigram and asked to produce any five-letter word that started with the given bigram and that met certain other restrictions detailed below. We were interested in whether early acquired words would tend to be produced more readily than later acquisitions.

Experiments 3 and 4 were exploratory studies intended to gather information on age-of-acquisition effects in the episodic memory tasks of free recall and recognition memory for recently presented words. Previous studies of age-of-acquisition effects have concentrated on lexical memory, but it would be worth knowing whether this variable also affects episodic memory. In order to construct a reasonable theory of age-of-acquisition effects, we first need to know where such effects appear. Experiments 3 and 4 were intended to add to the limited stock of information available on age effects.

All the experiments reported here involved the use of multiple-regression analyses to assess the relative contributions of age-of-acquisition scores in predicting various criterion measures (viz., picture naming speeds, frequency of producing target words to bigram cues, recall and recognition scores). In all cases, words were the units of analysis rather than subjects. This research strategy was adopted because the large number of potentially relevant variables that are correlated with age of acquisition (Gilhooly & Hay, 1977; Stratton et al., 1975) makes it impossible to carry out factorial experiments in which confounded variables are balanced

out or experimentally manipulated, while still retaining a reasonable number of words per condition. In common with a number of previous studies in this general area, then, we opted for the multiple-regression approach (Cohen & Cohen, 1975; Kerlinger & Pedhazur, 1973) in preference to the factorial design approach. In all cases the calculations were performed using the statistical package for the social sciences (SPSS) computer program (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975).

EXPERIMENT 1

This experiment was designed to examine the role of age of acquisition in picture naming when letter-order factors are taken into account. Also taken into account were codability, frequency, and some other commonly studied word attributes. Picturable words were selected as target names from a list of words for which measures were available on age, imagery, concreteness, familiarity, meaningfulness, and letter-order predictability (Gilhooly, 1978; Gilhooly & Hay, 1977). Times to produce the target names were analyzed using multiple-regression techniques to assess the relative contributions of the various name attributes and of picture codability.

Method

Subjects. The subjects in the picture naming task were 61 undergraduate student volunteers at Aberdeen University (28 males and 33 females). Subjects were paid £1 for their participation.

Stimulus materials. An initial set of 51 picturable five-letter words was taken from the Gilhooly and Hay (1977) list. A graphic technician made a black-and-white pen drawing to represent each of the initial words. Thus, for each picture there was a target name. The codability of the initial pictures was then checked with a group of 19 students drawn from the same population as the subjects used in the main task. These subjects were shown each picture for 20 sec and were asked to write down the first six names that they thought of for each picture. Using only the first names written down, the pictures for which there was less than 50% consensus with the target name were redrawn, and the revised versions were checked for codability with another group of 17 subjects.

The final selection consisted of 40 pictures ranging in degree of consensus with the target name from 41% to 100% (mean = 72.5%, SD = 14.9%). The target names of the pictures were as follows: ankle, apron, bench, brick, cairn, chair, cloak, clown, crowd, dunce, fairy, flame, flask, fruit, glove, gnome, graph, grave, house, judge, knife, match, medal, miner, mouth, pilot, plank, plant, pylon, rugby, scarf, scout, thief, thumb, tramp, tulip, watch, witch, woman, world.

Procedure. Subjects were tested individually. The pictures were presented in a tachistoscope, using a different random order for each subject. The sequence of events was as follows. When the subject said "ready," the experimenter caused a warning stimulus to appear to the subject in the tachistoscope, followed 2 sec later by one of the pictures. The picture remained visible for 4 sec. The subject was instructed to give a single word name for each picture as quickly as he or she could. A timer was started by the onset of the picture and was stopped by a voice key triggered by the subject's vocal response. Each subject was given five practice trials before the main series of 40 pictures was presented.

Word measures. Measures on each target name for age of acquisition, imagery, concreteness, familiarity, and meaningfulness were taken from Gilhooly and Hay's (1977) list. The Thorndike-Lorge (1944) frequency count was used as the source of frequency scores for each word. (Scores of AA were taken as 100 and scores of A as 50.)

The first letter-structure measure used was the total bigram rank (BR) score, allowing for bigram position and word length. This measure was obtained using Mayzner and Tresselt's (1965) tables in a way explained by Mendelsohn and O'Brien (1974). For five-letter words, the method involves drawing up a bigram frequency matrix of 20 rows and four columns for each word. In the matrix, the rows represent the 20 possible bigrams that can be formed from five different letters, and the columns represent the four bigram positions in a five-letter word. From Mayzner and Tresselt's tables, the frequency of each of the 20 possible bigrams in each of the four positions is obtained. Then the sum of the number of entries in the matrix that have higher frequencies is obtained across the four correct entries. The resulting number is the BR score for the word. This measure was found to be highly predictive of anagram difficulty (Mendelsohn, 1976; Mendelsohn & O'Brien, 1974), in that low BR facilitated solution. Presumably, because the greater the relative frequency of occurrence of correct bigrams within the set of bigrams that can be made from a given word, the more likely the subject is to generate a correct bigram as a memory probe. BR may be taken to be a measure of letter-order predictability.

The second measure is the number of nonzero entries in the bigram frequency matrix. This is a second measure of letter-order predictability. Mendelsohn (1976) found a separate negative effect on anagram solving for this variable (which he labeled "GTZERO"). The BR and GTZERO values for all the words in the Gilhooly and Hay (1977) list have now been tabulated (Gilhooly, 1978).

Results and Discussion

The average times to produce the target name for each picture were obtained (mean = 1,095.47 msec, SD = 272.41). Failures to produce the target name were not included in the calculations. The average time scores were positively skewed and were converted to speed scores by a reciprocal transformation. This effectively reduced the skew. The number of target responses to each picture (maximum = 61) was taken as a measure of codability. In view of the narrow range of imagery values exhibited by the picturable target names (mean = 5.92, SD = .42), this variable was not included in the analyses.

Table 2
Stepwise Multiple-Regression Analysis of Naming Speed Scores

Variable	Beta	F	df on Entry	R² Change	R
Codability	.57	17.99*	1,38	.32	.57
Age	−.37	8.14*	1,37	.12	.67
Concreteness	.20	2.58	1,36	.04	.69
Bigram Rank	−.16	1.55	1,35	.02	.71
GTZERO	−.10	.47	1,34	.01	.71
Familiarity	.04	.07	1,33	.00	.72
Meaningfulness	−.02	.02	1,32	.00	.72
Frequency	.02	.01	1,31	.00	.72

Note—The variables are listed from the top down in order of entry to the regression equation. The R^2 Change column indicates the increase in variance accounted for by adding each variable to the equation. The R column gives the multiple correlation when each variable is added.   *p < .01.

The product-moment correlations among the measures and some summary statistics are given in Table 1. From Table 1 it can be seen that codability and familiarity were positively correlated with speed of naming, while age and GTZERO were negatively correlated with naming speed.

Stepwise multiple regression (Table 2) indicated that the major effects were those of codability and age. The effects of familiarity, frequency, and letter-order predictability appear to have been largely redundant on age and codability. Age and codability together account for 44% of the criterion variance, while the remaining six variables account for only 7%. These results replicate previous findings of age-of-acquisition and codability effects in picture naming (Carroll & White, 1973; Lachman, 1973) and also indicate that letter-order predictability is not a major factor in this task, unlike anagram solving (Gilhooly & Johnson, 1978).

The magnitude of the age effect here (r = −.50 between age and naming speed) was markedly less than that obtained by Carroll and White (1973) (the corresponding r in their study was −.77). This difference may be due to the more limited range of the age variable explored in the present study. However, it may be noted that the correlations with naming speed obtained here

Table 1
Correlations Among Word Attributes, Codability, and Picture Naming Speed

	1	2	3	4	5	6	7	8	9
1. Speed	1.00	.57*	−.50*	.33*	.44*	.28	.13	−.24	−.37*
2. Codability		1.00	−.28	.16	.25	.43*	−.22	−.05	−.35*
3. Age			1.00	−.51*	−.56*	−.23	−.13	.27	.20
4. Frequency				1.00	.75*	.31	.10	−.46*	−.13
5. Familiarity					1.00	.32*	.31	−.36*	−.26
6. Meaningfulness						1.00	−.07	−.08	−.16
7. Concreteness							1.00	.07	.09
8. Bigram Rank								1.00	.51*
9. GTZERO									1.00
Mean	.01	45.58	2.70	42.58	4.61	5.47	5.89	53.13	39.13
SD	.002	14.40	.82	36.83	1.19	.86	.77	35.90	8.41

*Correlations significant at .05 level (two-tailed test).

for age, frequency, and familiarity variables were quite similar to those reported by Lachman (1973) in his study of individual picture naming data.

As a general explanation for age effects in picture naming and other retrieval tasks, it may be that the internal units or "logogens" (Morton, 1969) corresponding to early acquired words have lower thresholds than later acquired words, and so such words become available more readily to pictorial cues. The lowered thresholds of early words do not appear to be due to simple word frequency or to total life-span frequency, according to Carroll and White's (1977) analysis of the possible explanations for their picture naming data. Perhaps, though, early acquired words receive more frequent implicit arousal than their overt frequency would suggest. Since later acquired words are probably often learned through definitions in terms of early words, it could be that the occurrence of late words has some arousing effect on those earlier words involved in their definitions. Implicit arousals of this sort could lead to a lowering of thresholds for early acquired words. Whether this tentative hypothesis regarding the low thresholds of earlier words is ultimately upheld or not, the general threshold hypothesis is attractive in view of the range of retrieval tasks involving pictorial, semantic (Loftus & Suppes, 1972), and graphemic cues (Whaley, 1978) in which age effects have been found.

EXPERIMENT 2

This experiment focused on age effects in retrieval to a particular type of letter cue, the starting bigram. The starting bigram was chosen because anagram studies have indicated that the starting bigram is a useful cue (Dominowski, 1966), and studies concerned with word recall have demonstrated that the beginning portion of a word is easier to retrieve and serves as a better cue for the whole word than other parts (Brown & McNeill, 1966; Horowitz, White, & Atwood, 1968).

In the task used in the present experiment, subjects were given starting bigrams followed by three blanks (e.g., ST_ _ _) and were asked to fill in the blanks to make a five-letter noun meeting certain restrictions. For each of the 88 bigrams used, a target word had been selected from a list of five-letter words measured on imagery, concreteness, age of acquisition, familiarity, and meaningfulness (Gilhooly & Hay, 1977). The data were analyzed using multiple-regression methods in order to determine the relative contributions of the target-word attributes and of some objective measures in predicting hit rates per target.

Consideration of the task suggested that some objective measures could be useful predictors of hit rate. First, the number of words known to the subject that start with the required bigram should be relevant. This is the factor of "pool size" stressed by Duncan (1970). Essentially the same variable was labeled "vocabulary

size" by Murray (1975), and it played a significant role in his word production task. Measures relating to pool size were made for each bigram, and the resulting scores were labeled "bigram pool sizes." Both Duncan and Murray found that as pool size increased, so did the probability of producing at least one word from the pool. In those studies, any word from the pool was counted correct, while in the present study, the focus was on the probability of producing specified words from the pool. It could be that smaller pool sizes would produce higher hit rates in this task. Once the pool is delimited, the frequency of the target word and the frequencies of all other words in the pool (bigram pool frequency) could both be relevant since, as Duncan (1966) found, the most frequent words within the pool tend to be produced most readily. The relative frequency of the target word in its pool evidently depends on its own frequency and the total frequency of items in the pool, and so both types of measures were included in this study.

To summarize, Experiment 2 was designed to investigate the role of age of acquisition in retrieval to a specific type of letter cue (the starting bigram) when various other word attributes and frequency variables are taken into account. The experiment was intended to clarify interpretation of previous results (Gilhooly & Johnson, 1978) and to add to the information available on age-of-acquisition effects in retrieval tasks.

Method

Subjects. Forty-five student volunteers from the Aberdeen University student subject panel took part in the experiment. They were paid 50 pence.

Task. The subjects were given booklets containing 88 different starting bigrams, each followed by three blanks, and they were asked to fill in the blanks to make five-letter nouns (not plurals made by adding "s" or proper names). Subjects were asked to work as quickly as possible, and a time limit of 30 min was set for the whole task.

Target-word and cue measures. A single target was selected for each bigram from Gilhooly and Hay's (1977) list. If more than one word started with a given bigram, a random choice of target was made from the candidates. The measures of age of acquisition, imagery, concreteness, associative meaningfulness, and familiarity for each target word were also taken from the Gilhooly and Hay list.

The Thorndike-Lorge (1944) score was used as the objective frequency measure for each target word. (Scores of A were taken as 50 and AA scores as 100.) To obtain a measure of bigram pool size for each cue, it was assumed that the subjects would know most of the words in the Thorndike-Lorge list having frequencies of more than 1/million. Accordingly, a count was made of the number of such entries that started with each bigram cue. The total frequency of occurrence of five-letter words with each starting bigram (bigram pool frequency) was indexed by using Mayzner and Tresselt's (1965) table of bigram frequencies adjusted for word length and bigram position. The frequency of the bigram cue in the first position of five-letter words was taken as a measure of bigram pool frequency for each cue. The resulting score is the sum of frequencies of all five-letter words starting with that bigram in Mayzner and Tresselt's sample.

Each target word was given a hit score equal to the number of subjects who produced it in response to its bigram cue.

Table 3
Correlations Among Measures of Target Words, Cues, and Responses, Together with Summary Statistics

Measure	1	2	3	4	5	6	7	8	9	10
1. Root Hits	1.00	−.32*	.29*	.20	−.33*	.10	.22*	−.29*	.29*	−.39*
2. Total Responses		1.00	.05	−.05	−.19	.23*	.08	.44*	.24*	.33*
3. Imagery			1.00	.81*	−.63*	.35	.59*	.11	.29*	−.01
4. Concreteness				1.00	−.41*	.20	.47*	.09	.13	−.01
5. Age					1.00	−.61*	−.45*	−.25*	−.58*	.16
6. Familiarity						1.00	.41*	.34*	.67*	.25*
7. Meaningfulness							1.00	.07	.26*	.18
8. Root Bigram Pool Frequency								1.00	.32*	.42*
9. Root Frequency									1.00	.19
10. Root Bigram Pool Size										1.00
Mean	1.75	28.76	4.62	4.85	3.95	4.89	3.91	4.40	4.61	10.97
SD	1.19	9.87	1.19	1.24	1.30	1.04	1.12	2.93	2.97	4.88

Correlations significant at .05 level (two-tailed test).

There was some variation in the total number of responses to each bigram, and each target was also scored in terms of the total number of responses to its cue. The total number of responses could then be controlled for in the analyses.

Results

The hit scores, the Thorndike-Lorge (1944) frequencies, the bigram pool size scores, and the bigram pool frequency scores were positively skewed and so were subjected to square-root transformations. Table 3 presents the summary statistics and the (Pearson) correlations among the measures.

From the correlation matrix in Table 3, the number of hits was positively correlated with target-word imagery and meaningfulness but negatively correlated with target-word age of acquisition. Hit rate was also negatively correlated with total responses, bigram pool size, and bigram pool frequency, but positively correlated with target-word frequency. The intercorrelations among measures taken from Gilhooly and Hay's (1977) list matched the patterns obtained in that list as a whole.

In order to assess the relative contributions of the variables in predicting hit rates, a stepwise multiple regression was carried out. Since raw hit rate could be confounded with total number of responses, this variable was entered first in order to control for its effect on the criterion during the subsequent analysis. For the other independent variables, the order of entry was stepwise, so that the next variable to be entered was the one having the strongest correlation with the cirterion after the effects of total responses had been controlled for. The effect of each variable entered in the equation was partialled from the remainder and the next most important variable, after partialling, was added to the equation, and so on. Stepwise regression permits an assignment of percentage of criterion variance accounted for by each variable in the equation. The relevant figures are given in the R^2 column of Table 4 and indicate that total responses, target age of acquisition, bigram pool size, target frequency, and bigram pool frequency accounted for 10%, 16%, 12%, 4%, and 4%, respectively,

of the criterion variance when the present order of entry is followed.

The results of the regression analyses conflict with some expectations based on the correlation matrix of Table 3, in that imagery and meaningfulness, although having significant simple correlations with the criterion, did not make significant contributions to the regression equations. It appears that these variables were largely redundant and that their correlations with the criterion were due to their associations with age of acquisition.

Discussion

The results of Experiment 2 may be summarized as follows. Pool size had a negative relationship to hit rate. This is a reasonable result in that the more alternative words that start with a given bigram, other things being equal, the less likely the subject is to produce a given target rather than one of the alternatives. It may also be noted that pool size was positively related to total number of responses; this finding replicates those of Duncan (1970) and Murray (1975).

Target frequency was a significant predictor of hit rate. This suggests that within the bigram pool, the more frequent words are produced more readily (as Duncan, 1966, also found).

Of the variables not based on objective frequency

Table 4
Stepwise Multiple-Regression Analysis on Root Hit
Scores in Word Completion Task

Variable	Beta	F	df on Entry	R^2 Change	R
Total Responses	−.32	9.65**	1,86	.10	.32
Age	−.41	18.48**	1,85	.16	.51
Bigram Pool Size	−.38	17.06**	1,84	.12	.62
Frequency	.25	5.71*	1,83	.04	.65
Bigram Frequency	−.23	5.61*	1,82	.04	.68
Familiarity	−.17	2.21	1,81	.01	.69
Meaningfulness	.11	1.50	1,80	.01	.70
Concreteness	.01	.00	1,79	.00	.70
Imagery	.00	.00	1,78	.00	.70

*$p < .05$.   **$p < .01$.*

tables, only age of acquisition emerged as making a significant contribution to predicting hit rate. Early acquired target words are generally produced more often than later acquired targets. This result adds to the evidence from picture naming and category-instance naming (Carroll & White, 1973; Loftus & Suppes, 1972) that age of acquisition is an important variable in retrieval performance by adult subjects.

The relevance of these results to Gilhooly and Johnson's (1978) anagram experiment would appear to be as follows. Gilhooly and Johnson found no effect of rated age on anagram difficulty when measures of letter-order predictability were taken into account. They interpreted their results to mean that word retrieval to implicitly generated letter cues was not a major factor in their task as compared with letter manipulation processes. However, their results could be interpreted as suggesting that age of acquisition does not affect retrieval to letter cues. The present experiment supports the view that age of acquisition does affect retrieval to letter cues and lends support to Gilhooly and Johnson's original conclusions.

In discussing the results of Experiment 1, it was suggested that a logogen model (Morton, 1969) might be able to interpret age effects in picture naming. Such a model could also be invoked to explain the results of the present experiment. Making the reasonable assumption that logogens are sensitive to starting letters, a bigram cue would differentially activate those logogens matching the cue, and the first one to be activated above its threshold would make the first candidate word available for close examination. A logogen system could easily show pool size effects; the more logogens sensitive to a given cue, the less likely any specified one is to reach threshold first. Frequency effects are already accounted for in Morton's original scheme by a proposal that frequently aroused logogens gradually come to have lower thresholds than infrequently aroused logogens. If logogens corresponding to early acquired words also have reduced thresholds, the age effects would also be accounted for.

How could this be brought about? It could be argued that the total frequency with which a word has been experienced over the life span is the most important determinant of threshold levels. The life-span frequency of a word would be a multiplicative function of the time that the word has been known and its frequency of occurrence per n words. Thus, earlier acquired words, on average, would have higher life-span frequencies than later acquired words, and hence lower thresholds. Carroll and White (1973) considered a version of the life-span frequency argument in their study of picture naming. However, they found no multiplicative function of age and frequency that led to improved predictions of naming speed over linear functions. In the present experiment, a derived variable [root frequency X (8 − rated age of acquisition)] was computed for each word to give an

index of life-span frequency. (This measure was used because age of acquisition was rated from 1 for early acquired words to 7 for late acquired words.)

As in Carroll and White's (1973) study, the multiplicative function did not add anything to the separate contributions of rated age and root frequency. The simple correlation of the life-span frequency index and root hit rate was .30, but when rated age and root frequency were partialled out, the correlation fell to −.09. This suggests that age of acquisition and overt frequency do not interact in this task in the way suggested by the life-span frequency argument. As was suggested in discussing the picture naming data of Experiment 1, it may be that logogens corresponding to early words receive more frequent implicit arousals than their overt frequency suggests, thus leading to lowered thresholds for early acquired words.

Whatever the detailed explanation may be, the results of Experiment 2 support the view that words rated as early acquired are more readily produced to letter cues than words rated as later acquired. These results add to the accumulating evidence for age-of-acquisition effects in retrieval from lexical memory.

EXPERIMENT 3

Studies of the age-of-acquisiton variable have focused on retrieval from lexical memory. In order to examine the range of age-of-acquisition effects, the present experiment was designed to investigate the relative contributions of age, imagery, familiarity, and other word variables in predicting scores in the episodic memory task of free recall.

Method
Subjects. Thirty-nine student volunteers from the Aberdeen University student subject panel took part in the experiment. They were paid 50 pence.

Materials. Forty words were selected at random from Gilhooly and Hay's (1977) list of five-letter words measured on age of acquisition, imagery, concreteness, associative meaningfulness, and familiarity.

The Thorndike-Lorge (1944) scores were used as the objective frequency measures for each word. (Scores of A were taken as 50 and scores of AA as 100.)

The 40 words were randomly ordered in the presentation list. The same order of presentation was used for all subjects.

Procedure. The experiment was group administered. Subjects were informed that they would hear a list of words and at the end of the list, they were to write down as many words as they could remember from the list. There was a 2-sec interval between words. Recall was tested 10 sec after the last word had been presented. Subjects were given 10 min to complete the recall test.

Results and Discussion
Each word was scored in terms of the number of subjects who recalled it. To take account of serial position effects, each word was given a "serial position score" such that Words 1 and 40 were scored "20,"

Table 5
Correlations Among Word Attributes, Recall Scores, and List Position Indices

	1	2	3	4	5	6	7	8	9
1. Recall**	1.00	.47*	.16	−.26	.42*	.31*	.22	.14	.33*
2. Serial Position		1.00	.07	−.14	.12	.24	.16	.16	.23
3. Position			1.00	−.01	−.11	.12	.07	−.10	−.28
4. Age				1.00	−.59*	−.45*	−.20	−.65*	−.48*
5. Imagery					1.00	.82*	.28	.25	.26
6. Concreteness						1.00	.17	.08	.05
7. Meaningfulness							1.00	.26	.16
8. Familiarity								1.00	.68*
9. Frequency**									1.00
Mean	3.14	10.00	20.50	3.80	4.76	4.78	5.05	3.84	4.56
SD	1.14	5.86	11.69	1.05	1.06	1.26	.89	1.11	2.55

*Correlations significant at .05 level (two-tailed test).
**These variables have been subjected to square-root transformations to correct skews.

Words 2 and 39 were scored "19," and so on. This system meant that middle words had low serial position scores, while words at the extremes had high serial position scores. In addition, each word was given a simple list position scores ("1" for first, "2" for second, and so on). The frequency and the recall scores were positively skewed and were subjected to square-root transformations, which satisfactorily corrected the skews.

Table 5 presents the summary statistics and the Pearson correlations among the measures. The correlation matrix in Table 5 shows that recall scores were positively correlated with imagery, frequency, concreteness, and serial position scores. The results of a stepwise multiple regression of the recall scores as a function of the word measures indicate that serial position and word imagery were the major factors (Table 6). These variables account for 22% and 13% of the criterion variance, respectively, when the present order of entry is followed.

From both the simple correlation matrix and the regression analysis, it seems that age of acquisition had little or no effect on free recall scores. Serial position and imagery were the major determinants of free recall scores. The imagery result is consistent with most previous research (e.g., Paivio, 1971). These results indicate that age is not the effective variable underlying imagery in this task.

EXPERIMENT 4

In the final experiment, we sought to assess the relative contribution of word age of acquisition in performance on a recognition test for recently presented words. As in Experiment 3, this task involved memory for episodic information and was undertaken to obtain more data on the generality of age-of-acquisition effects beyond the lexical memory tasks that have been mainly studied hitherto.

Method

Subjects. Thirty student volunteers from the Aberdeen University student subject panel took part in the experiment. They were paid 50 pence.

Materials. One hundred to-be-recognized words were taken from Gilhooly and Hay's (1977) list of five-letter words measured on age of acquisition, imagery, concreteness, meaningfulness, and familiarity. Starting with the word "album," every other word in the list was used in the presentation list. The presentation list was prepared by randomly ordering the 100 words taken from the Gilhooly and Hay list.

The test list was constructed by taking another 100 words from the Gilhooly and Hay (1977) list. Each "old" word was paired at random with a "new" word, and subjects were asked to indicate on a response sheet which member of each pair had been on the presentation list. Old and new words were balanced for right/left position in the test pairs. The order of pairs on the test list was random.

A score was assigned to each word equal to the number of meanings recorded for it in *Collins English Gem Dictionary* (McLeod, 1963). Similar measures of semantic ambiguity have been found to be negatively related to recognition performance in previous research, presumably because of encoding variability effects (Morris, 1978).

Words were scored for frequency using the Thorndike-Lorge (1944) tables, in the same way as in Experiments 1-3.

Procedure. The experiment was group administered. Subjects were instructed that they would be shown a list of words, one at a time, on an overhead projector and that at the end of the list they would be given a recognition test. Each word was displayed for 2 sec, with a gap of 2 sec between words.

After the list had been presented, recognition test sheets were distributed; the subjects were allowed 15 min to work

Table 6
Stepwise Multiple-Regression Analysis of Recall Scores

Variable	Beta	F	df on Entry	R² Change	R
Serial Position	.47	10.92*	1,38	.22	.47
Imagery	.37	7.59*	1,37	.13	.60
Concreteness	−.31	1.69	1,36	.03	.62
Position	.27	1.69	1,35	.06	.67
Frequency	.16	1.22	1,34	.02	.68
Familiarity	−.28	2.81	1,33	.04	.71
Age	.05	.06	1,32	.00	.71
Meaningfulness	.01	.01	1,31	.00	.71

*p < .01.

Table 7
Correlations Among Word Attributes and Recognition Memory Scores

	1	2	3	4	5	6	7	8
1. Recognition	1.00	.41*	−.22*	−.10	−.20*	−.40*	−.45*	−.25*
2. Age		1.00	−.64*	−.46*	−.47*	−.71*	−.68*	−.35*
3. Imagery			1.00	.82*	.59*	.40*	.33*	.07
4. Concreteness				1.00	.47*	.29*	.18	−.03
5. Meaningfulness					1.00	.43*	.39*	.15
6. Familiarity						1.00	.76*	.47*
7. Frequency**							1.00	.61*
8. Dictionary Meanings**								1.00
Mean	24.78	3.83	4.68	5.06	5.04	3.85	4.15	1.63
SD	3.07	1.40	1.22	1.16	.89	1.16	3.21	.74

*Correlations significant at .05 level (two-tailed test).
**These variables have been subjected to square-root transformations to correct skews.

through these. They were instructed to make a response to each test pair and to guess if unsure.

Results and Discussion

Each "old" word was scored in terms of the number of subjects who indicated that they recognized it. The frequency and the number of dictionary meaning scores were positively skewed. The skews were satisfactorily corrected by means of square-root transformations.

Table 7 presents the summary statistics and the Pearson correlations among the measures. The correlation matrix in Table 7 shows that recognition scores were negatively correlated with frequency, familiarity, number of dictionary meanings, imagery, and meaningfulness, but were positively correlated with age of acquisition. The results of a stepwise multiple regression of the recognition scores (Table 8) indicate that frequency was the major factor. Frequency accounted for 21% of the criterion variance, while the remaining variables contributed only 2%.

It seems that age of acquisition had no independent effect on recognition scores in this study. The frequency result is consistent with previous research that has found rare words to be generally better recognized than common words.

The "new" or foil words were scored in terms of the number of subjects who falsely recognized them. However, analysis of these data uncovered no word variable that was significantly associated with probability of false recognition.

GENERAL DISCUSSION

Experiments 1 and 2 found that low age-of-acquisition words were more rapidly retrieved to pictorial cues and were more often produced to letter cues than were late age-of-acquisition words; Experiments 3 and 4 found no age-of-acquisition effects on recall or recognition of recently presented words. We now propose an interpretation of these results within a general view of long-term memory.

It seems reasonable to divide the contents of long-

term memory into various kinds of information. One important kind of information in memory is lexical information, that is, information about meaning, syntax, spelling, phonology, and articulation of the words in a person's vocabulary. Other important kinds of information that are presumably linked to lexicon entries are encyclopedic and episodic information. For example, corresponding to the word "dog" would be basic lexical information about the word itself, encyclopedic information about dogs in general, and episodic information about personal experiences with dogs. This division incorporates distinctions made by Lachman (1973) and Tulving (1972). Tulving distinguished between semantic and episodic memory, but the arguments put forward by Lachman lead to the subdivision of semantic memory into lexical and encyclopedic components, both of which represent information different from the personal contextual-temporal information incorporated in episodic memory.

Lexical memory is involved in all verbal processes, but it plays a more important role in some tasks than in others. The tasks in Experiments 1 and 2 presumably involved mainly retrieval from lexical memory with little involvement of episodic information; the tasks in Experiments 3 and 4 drew heavily on episodic memory. One way to conceptualize the process of lexical retrieval is to invoke a logogen-like system. This kind of system has been found useful in handling data from tachisto-

Table 8
Stepwise Multiple-Regression Analysis of
Recognition Memory Scores

Variable	Beta	F	df on Entry	R² Change	R
Frequency	−.45	25.07*	1,97	.21	.45
Age	.18	2.15	1,96	.02	.47
Concreteness	.06	.34	1,95	.00	.48
Imagery	−.10	.34	1,94	.00	.48
Familiarity	−.06	.18	1,93	.00	.48
Dictionary Meanings	.04	.12	1,92	.00	.48
Meaningfulness	.02	.03	1,91	.00	.48

*$p < .01$.

scopic word recognition (Morton, 1969) and lexical decision tasks (Meyer & Schvaneveldt, 1976), and it would seem capable of useful extension to the tasks studied here.

The main assumption of the logogen approach is that corresponding to each entry in lexical memory, there is a theoretical device (i.e., a logogen) that accepts sensory and contextual information relevant to that particular word. When more than a threshold amount of information has accumulated in any logogen, the word associated with the logogen becomes available for overt production. In Morton's (1969) original proposal, it was assumed that following each occasion on which a word becomes available, the threshold of the relevant logogen is permanently reduced to a small extent. So, over time, the logogens of words that are produced or received with high frequency come to have low thresholds.

The results of Experiments 1 and 2, together with other data in the literature, suggest that the logogens corresponding to early age-of-acquisition words also have low thresholds. Confounding of age with simple frequency or with life-span frequency does not explain the age effect (Carroll & White, 1973, Experiment 2). We have proposed that the advantage of early age-of-acquisition words in lexical retrieval tasks may be due to a lowering of logogen thresholds resulting from priming by the occurrence of later words that are semantically related to early acquired words. In support of this proposal, it may be noted that later acquired words are probably often learned by definitions in terms of early words, and so, it is plausible to suppose that the occurrence of a late word could have an arousing effect on the logogens of earlier words involved in its definition. Thus, the occurrence of a later acquired word could make available earlier acquired words and so lead to a reduction in the logogen thresholds of the earlier words. In this way, the thresholds of early acquired words would be lower than would be predicted from simple or life-span word frequency. Whatever the detailed mechanism, a general threshold hypothesis is consistent with the effects of age of acquisition in the lexical memory tasks studied in Experiments 1 and 2.

No effects were found for age of acquisition in the present studies of episodic memory tasks (Experiments 3 and 4). Although lexical memory must be involved in episodic tasks, other more influential factors apparently come into play. First, we consider processes that may be involved in the free recall task. In the study phase of free recall, items are presented one at a time. As each word is presented, the lexicon is accessed and a meaning for the word is retrieved. Encyclopedic and episodic information associated with the word may also be retrieved, and all the varieties of information may be used to try to link each incoming word to its predecessors, so as to form a useful retrieval scheme for the recall phase. It would seem that differences in item

accessibility in the lexicon (reflecting logogen threshold differences) have only a small effect on the times left on each trial to engage in mnemonic processing, and so have little effect on final recall. The results of Experiment 3 supported a number of previous studies (Paivio, 1971) in finding that word imagery was a major determinant in free recall. This is presumably because high-imagery words can be more readily incorporated into a developing retrieval plan than low-imagery words. Neither word age of acquisition nor frequency played a major role in predicting our free recall data.

Lexical memory must also be involved in episodic recognition tasks using words. During the study phase, it seems likely that when a word is presented a selection of attributes associated with that word (incorporating lexical and other information) is encoded into episodic memory together with a tag indicating that this attribute complex is on the current list. Numerous studies of encoding effects indicate that it is not the word that is stored during the study phase, but some semantic interpretation of the word (Baddeley, 1976, Chapter 11). During the test phase, the test words can be used to generate attributes from lexical memory that can then be compared with encoded information in episodic memory to decide which test items are old. The greater scope for encoding variability between study and test for high-frequency words, as compared with low-frequency words, may have been the main factor in Experiment 4 underlying the correlation between rarity and recognition scores. A number of authors (e.g., Morris, 1978) have proposed such an explanation for the rarity effect in word recognition memory. Although age-related differences in accessibility (which may reflect logogen threshold differences) might have affected times to retrieve word meanings during the study phase, the effectiveness with which late acquired words were encoded was not impaired. Indeed, if anything, late acquired and low-frequency words were recognized better than early acquired high-frequency words. Such words tend to have fewer meanings and associations available, which reduces encoding variability and thus enhances recognition.

To summarize, the results of these studies indicate that early age of acquisition facilitates retrieval from lexical memory but has little or no effect in episodic memory tasks.

REFERENCES

BADDELEY, A. D. *The psychology of memory.* New York: Basic Books, 1976.

BUTTERFIELD, G. B., & BUTTERFIELD, E. C. Lexical codability and age. *Journal of Verbal Learning and Verbal Behavior,* 1977, 16, 113-118.

BROWN, R. W., & McNEILL, D. The "tip-of-the-tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior,* 1966, 5, 325-337.

CARROLL, J. B., & WHITE, M. N. Word frequency and age of

acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology*, 1973, **25**, 85-95.

COHEN, J., & COHEN, P. *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, N.J: Erlbaum, 1975.

DOMINOWSKI, R. L. Anagram solving as a function of letter moves. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 107-111.

DUNCAN, C. P. Effect of word frequency on thinking of a word. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 434-440.

DUNCAN, C. P. Thinking of a word under different retrieval constraints. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 356-361.

GILHOOLY, K. J. Bigram statistics for 205 five-letter words having single-solution anagrams. *Behavior Research Methods & Instrumentation*, 1978, **10**, 389-392.

GILHOOLY, K. J., & HAY, D. Imagery, concreteness, age-of-acquisition, familiarity, and meaningfulness values for 205 five-letter words having single-solution anagrams. *Behavior Research Methods & Instrumentation*, 1977, **9**, 12-17.

GILHOOLY, K. J., & JOHNSON, C. E. Effects of solution word attributes on anagram difficulty: A regression analysis. *Quarterly Journal of Experimental Psychology*, 1978, **30**, 57-70.

HOROWITZ, L. M., WHITE, M. A., & ATWOOD, D. W. Word fragments as aids to recall: The organization of a word. *Journal of Experimental Psychology*, 1968, **76**, 225-232.

KERLINGER, F. N., & PEDHAZUR, E. J. *Multiple regression in behavioral research*. New York: Holt, Rinehart, & Winston, 1973.

KINSBOURNE, M., & EVENS, N. Is the word frequency effect on recognition threshold a function of transitional probability? *Psychological Reports*, 1970, **27**, 143-145.

LACHMAN, R. Uncertainty effects on time to access the internal lexicon. *Journal of Experimental Psychology*, 1973, **99**, 199-208.

LACHMAN, R., SHAFFER, J. P., & HENNRIKUS, D. Language and cognition: Effects of stimulus codability, name-word frequency, and age of acquisition on lexical reaction time. *Journal of Verbal Learning and Verbal Behavior*, 1974, **13**, 613-625.

LANDAUER, T. K., & STREETER, L. A. Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 119-131.

LOFTUS, E. F., & SUPPES, P. Structural variables that determine the speed of retrieving words from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 770-777.

MAYZNER, M. A., & TRESSELT, M. E. Tables of single-letter bigram frequency counts for various word length and letter position combinations. *Psychonomic Monograph Supplement*, 1965, **1**, 13-31.

MCLEOD, W. T. (Ed.). *Collins English gem dictionary*. London: Collins, 1963.

MENDELSOHN, G. A. A hypothesis approach to the solution of anagrams. *Memory & Cognition*, 1976, **4**, 637-642.

MENDELSOHN, G. A., & O'BRIEN, A. T. The solution of anagrams: A reexamination of the effects of transition letter probabilities, letter moves, and word frequency on anagram difficulty. *Memory & Cognition*, 1974, **2**, 566-574.

MEYER, D. E., & SCHVANEVELDT, R. W. Meaning, memory structure, and mental processes. In C. N. Cofer (Ed.), *The structure of human memory*. San Francisco: Freeman, 1976.

MORRIS, P. E. Frequency and imagery in word recognition: Further evidence for an attribute model. *British Journal of Psychology*, 1978, **69**, 69-75.

MORTON, J. Interaction of information in word recognition. *Psychological Review*, 1969, **76**, 165-178.

MURRAY, D. J. Graphemically cued retrieval of words from long-term memory. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, **104**, 65-70.

NIE, N. H., HULL, C. H., JENKINS, J. G., STEINBRENNER, K., & BENT, D. H. *SPSS: Statistical package for the social sciences*. New York: McGraw-Hill, 1975.

PAIVIO, A. *Imagery and verbal processes*. New York: Holt, Rinehart, & Winston, 1971.

STRATTON, R. P., JACOBUS, A. J., & BRINLEY, B. Age-of-acquisition, imagery, familiarity, and meaningfulness norms for 543 words. *Behavior Research Methods & Instrumentation*, 1975, **7**, 1-6.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teacher's College Press, 1944.

TULVING, E. Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972.

WHALEY, C. P. Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, 1978, **17**, 143-154.