

# Associative relatedness vs synonymy in the false-recognition effect

DONALD H. KAUSLER and ANITA V. SETTLE  
*University of Missouri, Columbia, Mo. 65201*

The false-recognition effect was demonstrated for test words that were associatively related to prior stimulus words. The extent of the effect was seemingly independent of the presence/absence of a semantic relationship between test and stimulus words. In addition, the extent of the effect covaried positively with the degree of associative relatedness between experimental test words and prior stimulus words when false recognition was measured by a relative score derived from confidence ratings. The results favored the implicit associative response hypothesis.

False recognitions of new test words as being "old" occur more frequently when the new words are related to prior study words than when they are unrelated (e.g., Anisfeld & Knapp, 1968; Grossman & Eagle, 1970; Underwood, 1965). Underwood (1965) attributed the false-recognition effect to the elicitation of new words as implicit associative responses (IARs) to their related prior words in the study list. That is, a given new test word is assumed to have had an implicit occurrence before it appears formally for the first time as a test item. IARs add frequency units to new items, thereby increasing the difficulty of "old" vs "new" discriminations that are presumed to be mediated by frequency cues. Alternatively, Anisfeld & Knapp (1968) attributed the false-recognition effect to a feature-tagging principle. Semantic and sensory features of study words are hypothesized to be tagged for recency of occurrence at the time of their initial presentation. Later recognition of test words as "old" items is contingent upon the extent to which these words possess features that were tagged earlier for recency of occurrence. Since related new words overlap in featural content with prior study words, their rate of misidentification as "old" items exceeds that found for unrelated new words (i.e., the false recognition effect).

One means of contrasting the IAR and feature-tagging hypotheses is to compare the false-recognition rates for sets of new test items that are equated in terms of their normative associative relatedness to prior study items, but differ in terms of their degree of synonymy. That is, the comparison is between associatively equated test words that are synonyms of study words, in one case, and are neither synonyms nor antonyms, in the other case. The IAR hypothesis predicts no difference in false-recognition rates for the two cases, in that they represent equal associative strengths. On the other hand, the feature-tagging hypothesis predicts a higher false-recognition rate for synonyms than for nonsynonyms. This prediction follows from the greater

featural overlap between study items and synonymous test items than between study items and nonsynonymous test items.

Previous comparisons have yielded ambiguous results. In agreement with the feature-tagging hypothesis, Fillenbaum (1969) reported a higher false-recognition rate (17%) for synonyms than for associatively equated nonsynonyms (12%). However, there were no control test items that provided a baseline for evaluating false-recognition rates in the absence of either associative relatedness or synonymy. Ideally, the control item for any given experimental item matches the latter in such attributes as frequency count, number of syllables, and grammatical form class. It is conceivable that Fillenbaum's synonyms differed from his nonsynonyms in terms of one or more of these other attributes (e.g., grammatical form class), and that at least part of the disparity in false-recognition rates resulted from these differences. In agreement with the IAR hypothesis, Grossman & Eagle (1970) found nearly equal absolute false-recognition rates for synonyms (20.3%) and nonsynonyms (21.1%). Grossman and Eagle also included a control test item for each of their experimental test items. Both synonyms and nonsynonyms yielded significantly greater false-recognition rates than do their appropriate control items. In addition, the difference between experimental and control rates was somewhat greater for synonyms (8.6%) than for nonsynonyms (4.8%)—a finding expected on the basis of the feature-tagging hypothesis. However, since Grossman and Eagle did not test the Synonym-Nonsynonym by Experimental-Control interaction effect, the statistical significance of their observed difference in rates is uncertain.

The present study attempted to reduce the ambiguity by replicating, in part, Grossman and Eagle's study with a design permitting a direct test of the critical Synonym-Nonsynonym by Experimental-Control interaction effect. In addition, the present study extended the earlier one by including a confidence rating for each "old" vs "new" test-item decision. Hall (1969) has demonstrated convincingly that such ratings provide a more sensitive measure of the false-recognition effect than do error rates. Of special interest in this study was the covariation between confidence ratings and the degree of normative associative strength between experimental test items and their prior study items. Using absolute error rates, Grossman and Eagle failed to find a significant positive correlation (in fact, if anything, it was slightly negative) between the magnitude of the false-recognition rate and the degree of associative strength. This is, of course, contrary to the

**Table 1**  
**Summary Statistics for Recognition Errors**  
**and Confidence Ratings**

Semantic Class	Errors		Ratings	
	Experimental	Control	Experimental	Control
Related				
Mean	4.25	2.75	7.55	8.24
SD	3.01	2.25	1.24	.99
Unrelated				
Mean	3.25	1.50	8.02	8.48
SD	2.57	1.85	.86	.97

implications of the IAR hypothesis, and it also conflicts with the finding of a substantial positive covariation in children (e.g., Cramer, 1972). However, it is conceivable that the absence of a positive correlation in adults merely reflects the insensitivity of absolute error rates.

## METHOD

### Subjects and Design

The Ss were 40 students in general psychology classes at the University of Missouri, all of whom were naive with respect to prior participation in verbal learning experiments and all of whom earned extra course credit for their participation in this experiment. They were assigned in randomized blocks to two groups (Groups I and II), which differed only with respect to specific word content. For both groups, there were 8 critical stimulus (CS) words embedded within the 72 items presented during a single study phase. Each CS word was associatively related to a test word embedded within the 66 items presented, during a single test phase. Of the 8 CS words per group, 4 were related semantically to their later test words via synonymy and 4 were unrelated (i.e., they were neither synonyms nor antonyms of later test words). Different relationships between other sets of study-test items were included for purposes irrelevant to the present study. These inclusions were part of a pilot investigation that will not be described here. However, it should be noted that the 64 non-CS words in the study list were carefully selected for absence of relatedness with CS and critical test words.

### Lists and Procedure

The 8 semantically related and the 8 semantically unrelated CS-test word combinations were selected from the 9 combinations of each category employed earlier by Grossman & Eagle (1970).<sup>1</sup> The two categories represented in the present study were approximately equal in degree of normative associative relatedness (range = 23.0%-35.2%, mean = 27.6%, for the semantically related combinations; range = 20.1%-41.2%, mean 27.9%, for the semantically unrelated combinations). Two sets (Sets I and II) of CS-test word combinations were constructed by randomly assigning 4 semantically related and 4 semantically unrelated combinations to each set. Groups I and II received Sets I and II, respectively. Of the 66 test words, 8 were associates of the CS words, 8 were control words (the same words employed earlier by Grossman and Eagle; each control word matched one of the associates in terms of frequency count, grammatical form class, and number of syllables), 30 were old items from the study phase, and 20 were new filler items.

Study items were presented individually by means of a Carousel slide projector. Each item was exposed on a screen for 3 sec, during which time S was required to pronounce the word twice. Four different input orders were employed for each group (5 Ss per order) as a control for possible serial position effects. The test phase followed immediately after the study phase. Each

S received a booklet containing 66 pages. At the top of each page was a test item. S was instructed to read each test word and to check, in the appropriate space below it, whether it was an "old" or "new" word. In addition, S rated the confidence of his "old" vs "new" decision on a 5-point scale, ranging from "don't know" to "positive." There was a different random order of pages in the booklet for each S. However, the first four test items were always new filler items. Performance on the test phase was self-paced.

## RESULTS AND DISCUSSION

An error score, defined as the number of Ss falsely identifying a test item as "old" (maximum score = 20), was determined separately for each experimental word and for each control counterpart. Summary statistics for semantically related and semantically unrelated experimental and control words (collapsed over Sets I and II) are given in Table 1. These error scores were subjected to a 2 by 2 by 2 mixed analysis of variance in which the individual experimental and control words served as data units (total N = 32). Set I vs Set II and related vs unrelated (semantically) functioned as between-group variables, and experimental vs control as a within-group variable. Only the experimental vs control main effect (i.e., false recognition) attained statistical significance,  $F(1,12) = 12.03$ ,  $p < .005$ . Most importantly, the Experimental-Control by Related-Unrelated interaction effect fell far short of statistical significance,  $F(1,12) < 1$ . Within the limits imposed by the acceptance of the null hypothesis, the present results imply that the false-recognition effect is no greater for semantically related experimental words than it is for unrelated experimental words when the degree of associative relatedness to CS words is equated for the two classes.

Confidence ratings were scored for each experimental and control word by assigning the values 1-10 to the continuum, "new word, positive" (1)—"old word, positive" (10). Summary statistics for semantically related and semantically unrelated experimental and control words (collapsed over Sets I and II) are included in Table 1. A 2 by 2 by 2 analysis of variance for these confidence rating scores revealed that only the experimental vs control main effect reached statistical significance,  $F(1,12) = 6.77$ ,  $p < .025$ . The Experimental-Control by Related-Unrelated interaction effect again fell far short of statistical significance,  $F(1,12) < 1$ . The results obtained with confidence ratings thus replicated those obtained with error rates.

In agreement with Grossman & Eagle (1970), the correlation between absolute error rate for experimental words and the degree of associative relatedness to CS words was low ( $\rho = .07$ ) and nonsignificant. The correlation between absolute confidence ratings for experimental words and the degree of associative relatedness was also low ( $\rho = .03$ ) and nonsignificant. For purposes of conducting a more sensitive test of the covariation between the magnitude of the false-recognition effect and the degree of associative

relatedness with CS words, a difference score was computed for each experimental word. The difference score for a given experimental word was found by subtracting the mean confidence rating for its control counterpart (i.e., the control word matched for frequency, etc.) from the word's own mean confidence rating. The correlation between these difference scores and degree of associative relatedness approached statistical significance,  $\rho(16) = .44$ ,  $p < .10$ . While moderate, the correlation is nevertheless impressive in view of the modest range of variation in degree of associative relatedness between experimental and CS words.

In general, the present results favor the IAR hypothesis. That is, the presence of associative relatedness alone is seemingly sufficient cause for generating the false-recognition effect. However, the possibility remains that associatively related words that are neither synonyms nor antonyms may, nevertheless, be related semantically via other, unmeasured, shared semantic features. Analyses of these semantic features, and the determination of their role in mediating the

false-recognition effect, remain problems for future research.

## REFERENCES

- Anisfeld, M., & Knapp, M. E. Association, synonymy, and directionality in false recognition. *Journal of Experimental Psychology*, 1968, 77, 171-179.
- Cramer, P. Evidence for a developmental shift in the basis for memory organization. Paper presented at the annual meeting, Psychonomic Society, St. Louis, November 1972.
- Fillenbaum, S. Words as feature complexes: False recognition of antonyms and synonyms. *Journal of Experimental Psychology*, 1969, 82, 400-402.
- Grossman, L., & Eagle, M. Synonymy, antonymy, and association in false recognition. *Journal of Experimental Psychology*, 1970, 83, 244-248.
- Hall, J. W. Effects of IAR occurrence during learning on confidence in judgments during recognition. *Journal of Experimental Psychology*, 1969, 79, 578-580.
- Underwood, B. J. False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, 1965, 70, 122-129.

## NOTE

1. We are grateful to Grossman and Eagle for making their word lists available to us.

(Received for publication May 25, 1973.)

# The influence of time course variables on REM sleep

H. W. AGNEW, JR., and W. B. WEBB

*University of Florida, Gainesville, Fla. 32601*

This paper studies the effect of age, length of prior wakefulness, length of sleep, and a circadian influence on REM sleep. Age is a major determinant of REM amount up to the early teens, but has little effect thereafter. Length of prior wakefulness has little or no effect on REM amount. Length of sleep has a major influence on REM amount, with longer sleep times producing greater amounts of this type of sleep. There is a strong circadian effect on REM, with early night sleep periods producing little REM and early morning periods a relatively great amount of REM.

In an earlier report (Webb, Agnew, 1972), we reviewed and analyzed the responsiveness of Stage 4 sleep (deep sleep) to changes in four variables: age, length of the sleep period, length of prior wakefulness, and the onset time of the sleep period (a circadian effect). Because our data were drawn from studies not originally designed specifically to study these influences, the data on prior wakefulness was confined to the first 3 h of sleep and the data on circadian effects were limited in onset points. Nevertheless, since Stage 4 sleep is generally distributed within the first 3 h of sleep we felt confident of our generalizations.

This paper reviews the same data relative to changes in four variables and REM sleep. Stage REM is associated with rapid eye movements and a Stage 1 electroencephalogram (EEG). This stage of sleep was first recognized by Aserensky & Kleitman (1953) and is associated with visual dreaming in humans (Dement & Kleitman, 1957). There are numerous reports that when Ss are deprived of REM sleep they tend to recover it on subsequent nights in a manner which would suggest that it represents a basic need state (Dement, 1960). Since, as will be noted below, this stage of sleep is predominant after the first 3 h of sleep, the data relative to prior wakefulness is severely limited. However, it is likely to be some time before the needed data on age and length of sleep are available, but the nature of the influence of prior wakefulness on REM in the first 3 h of sleep show interesting relations to Stage 4, and we feel that this limited interim report is appropriate.

The data for the present report was gleaned over the course of 10 years' research on human sleep patterns. Our procedures for obtaining EEG sleep records and for sleep stage scoring are described in a comprehensive report (Agnew & Webb, 1972). Essentially, a 1-min epoch of EEG is scored as Stage REM if rapid eye movements occur contiguously with a Stage 1 EEG. The