Journal of Experimental Psychology:
Learning, Memory, and Cognition
2004, Vol. 30, No. 2, 319–331

0278-7393/04/$12.00 DOI: 10.1037/0278-7393.30.2.319

Modeling the Effects of Repetitions, Similarity, and Normative Word Frequency on Old–New Recognition and Judgments of Frequency

Kenneth J. Malmberg
Iowa State University of Science and Technology

Jocelyn E. Holden and Richard M. Shiffrin
Indiana University Bloomington

Judgments of frequency for targets (old items) and foils (similar; dissimilar) steadily increase as the number of times a target is studied increases, but discrimination of targets from similar foils does not steadily improve, a phenomenon termed *registration without learning* (D. L. Hintzman & T. Curran, 1995; D. L. Hintzman, T. Curran, & B. Oppy, 1992). The present experiment explores this phenomenon with words of differing normative word frequency. The retrieving-effectively-from-memory model (REM; R. M. Shiffrin & M. Steyvers, 1997, 1998) predicts that low-frequency words will be better recognized than high-frequency words because low-frequency words have more distinctive memory representations. A corollary of this assumption predicts that the typical recognition word-frequency effect will be disrupted when similar foils are tested. These predictions were confirmed, but to fit both the recognition and the judgment-of-frequency data, the authors used a "dual-process" extension of the REM model.

Explicit memory usually improves as the number of times that an item is studied increases (e.g., Crowder, 1976, for a review). However, Hintzman, Curran, and Oppy (1992; see also Hintzman & Curran, 1995; Sheffert & Shiffrin, 2003) found a noteworthy exception. In their experiments, participants studied items up to 25 times at spaced intervals, and a given item was always studied in the same form (e.g., always in the same plurality).[1] Test items included old items (targets; e.g., *TOAD*), new items (dissimilar foils; e.g., *CAKE*), and items that are semantically and perceptually very similar to targets (similar foils; e.g., *TOADS*). At test, participants gave a judgment of frequency (JOF), indicating how many times the test item was studied. A JOF > 0 indicates that the item was studied at least once, a JOF of 0 indicates that the item was not studied, and participants are explicitly instructed to respond "0" to similar foils. In these experiments, mean JOFs for targets and similar foils were found to be a strictly positive function of the number of target presentations, but discrimination of targets from similar foils did not improve after the first two or three presentations. Hintzman et al. (1992) termed this finding *registration without learning* because the steady increase in JOFs indicates that successive presentations are "registered" in memory but the failure to find a steady increase in discrimination indicates that the features critical for discriminating targets and similar foils are not "learned" after the first two or three presentations.

The present research has two objectives. One objective is to empirically test an account of normative word-frequency effects in paradigms like the ones under discussion. However, our results will also bear on several issues not involving word frequency. For example, why do only the initial two or three presentations of an item improve discrimination of targets from similar foils? Why are old–new recognition judgments and JOFs affected differently by later repetitions? To address these issues, we developed a formal model, and we therefore begin with a discussion of some accounts of the registration-without-learning findings.

Accounts of Registration Without Learning

Discrimination is usually defined as an increasing function of the probability of correctly classifying a studied item (i.e., hit rate [HR]) and a decreasing function of the probability of incorrectly classifying an unstudied item (i.e., false-alarm rate [FAR]). Thus, the finding that HRs for targets and FARs for similar foils are not affected by later repetitions is consistent with the conclusion that discrimination of targets and similar foils does not improve with later repetitions (e.g., Hintzman & Curran, 1995). Explaining the registration-without-learning findings is not easy within the framework of the most common quantitative models of recognition memory (e.g., McClelland & Chappell, 1998); such models posit a global matching of the test item to memory, the generation of a "familiarity" value, and a decision based on comparison of the familiarity to a criterion. If the decision system does not take into account the fact that some items very similar to studied items are

Kenneth J. Malmberg, Department of Psychology, Iowa State University of Science and Technology; Jocelyn E. Holden and Richard M. Shiffrin, Department of Psychology, Indiana University Bloomington.

This research was supported in part by National Institute of Mental Health (NIMH) National Research Service Award Postdoctoral Fellowship MH12643 to Kenneth J. Malmberg and NIMH Grants 12717 and 63993 to Richard M. Shiffrin.

Correspondence concerning this article should be addressed to Kenneth J. Malmberg, Department of Psychology, Iowa State University of Science and Technology, W112 Lagomarcino Hall, Ames, IA 50011, or to Richard M. Shiffrin, Department of Psychology, Indiana University, Bloomington, IN 47405. E-mail: malmberg@iastate.edu or shiffrin@indiana.edu

[1] Hintzman and Curran's (1995) findings generalize from words varying in plurality to photographs varying in left–right orientation. In the present studies, we consider only verbal stimulus materials.

tested, then most such models predict that the number of times an item is studied increases the probability of responding of “old” to both targets and similar foils (Hintzman et al., 1992; C. M. Jones & Heit, 1993; Shiffrin, Huber, & Marinelli, 1995; see Clark & Gronlund, 1996, for a review of global-memory models) because they do not have mechanisms that provide any obvious reason why learning of the features that discriminate between targets and dissimilar foils would increase across repetitions when learning the features that discriminate targets from similar foils would not. Hence, these global-matching models can predict the patterns of JOFs but not the pattern of HRs and FARs.

The different patterns of JOFs and old–new recognition data suggest that different mechanisms might be responsible for performing the tasks. For instance, Hintzman et al. (1992; also Hintzman, 1988) posited that JOFs depended on something akin to familiarity (summed global matching in some form), whereas old–new discrimination required a mechanism for recalling the critical feature(s) that discriminate between targets and similar foils. If so, the outcome of a recall-like process can be used to correctly reject similar foils by comparing the recalled information to test item, and if additional repetitions improve the recall of critical features, then the increased rejection of similar foils based on recalling those features will offset or attenuate the increase in their familiarity (cf. Jacoby, 1991). Models like these are often called “dual-process” models, and they share the assumption that repetitions produce storage of features that improve discrimination of targets from similar foils, and do so throughout the course of study. Hence, this alternative may be called the *registration-with-learning* hypothesis.

Some evidence consistent with the registration-with-learning hypothesis comes from experiments in which discrimination between targets and similar foils does improve with later repetitions (e.g., Hintzman et al., 1992, Experiments 2 and 4). For example, Rotello, MacMillan, and Van Tessel (2000) tested recognition memory with foils that differed from targets only in plurality and with new foils. In addition, they conducted an analysis of receiver operating characteristic (ROC) functions because some dual-process models predict that the ROC will become more linear as the recall contributes more to recognition in the similar foil condition (Yonelinas, 1994; but see Malmberg, 2002), and this prediction was confirmed. Using a similar procedure, Curran (2000; also see Curran & Cleary, 2003) used event-related potential technology to measure brain activity during old–new recognition judgments, and he observed different patterns of brain activity depending on the type of test item: The parietal 400–800 signal, which Curran assumed was a marker for recollection (cf. Rugg, Cox, Doyle, & Wells, 1998), was greater when a hit was made than when a false alarm was made to a similar foil. Kelley and Wixted (2001) and T. C. Jones and Jacoby (2001) used an associative recognition design in which intact and rearranged word pairs were discriminated. Both sets of experiments produced similar results: Repetitions increase HRs, but they have little or no effect on FARs for rearranged word pairs, resulting in an improvement in the ability to discriminate targets from foils.

All these research groups interpreted their findings to reflect that a recall-like process and a familiarity-based process served together to decrease and increase FARs, respectively: The familiarity of a test item and the result of a recall-process respectively indicate that an item is very familiar and that it was not studied because critical features are either recalled or not recalled. To this point, however, the processes involved in performing this task have not been formally described. The model that we later describe accomplishes this goal.

Normative Word-Frequency and Item-Similarity Effects

Another focus of the present research involves the relation of foil similarity to normative word frequency. When normative word frequency is varied, discrimination is better for low-frequency (LF) items, a phenomenon known as the word-frequency effect (WFE; Shepard, 1967). Importantly, the HR is greater for LF words than for high-frequency (HF) words, and the FAR is less for LF words than for HF words, a phenomenon known as a “mirror effect” (Glanzer & Adams, 1985; but see Wixted, 1992). For studies like that of Hintzman et al. (1992), a mirror effect is defined as $P(\text{JOF} > 0)$ less for LF than HF foils, and $P(\text{JOF} > 0)$ greater for LF than for HF targets.

In the retrieving effectively from memory (REM) model of recognition memory (Shiffrin & Steyvers, 1997, 1998), memory traces and retrieval cues are vectors of features. Each studied item produces a “noisy” episodic vector. The test vector is compared feature by feature to each such episodic vector; the comparison produces a likelihood ratio for each stored vector. The average of the likelihood ratios gives the odds that the test item is old, and the default decision rule is therefore “respond old if the odds is greater than 1.0.”[2]

In REM (Malmberg & Murnane, 2002; Malmberg, Steyvers, Stephens, & Shiffrin, 2002; Malmberg, Zeelenberg, & Shiffrin, 2004; Shiffrin & Steyvers, 1997, 1998), features representing past events vary in their environmental frequency, or base rate. Rare features are relatively more “diagnostic,” and therefore a match between a rare probe feature and a corresponding feature in memory provides more evidence in favor of the probe being “old” because rare features are unlikely to be encountered by chance alone. Thus, a match of a rare feature contributes more evidence to the likelihood ratio than a match of a common feature. REM accounts for the recognition WFE by assuming that the memory representations of LF words tend to be made up of less common and therefore more diagnostic features than the memory representations of HF words. As a result, LF targets tend to match their own memory traces to a greater degree than HF targets, but HF foils tend to match the traces of *other* words to a greater degree than LF foils because HF words tend to share relatively common features and therefore match better by chance. The mirror effect is predicted by REM through an interaction of these factors with assumptions concerning how matching is calculated and how decisions are made. In particular, it is assumed that calculations of matching are carried out not on the basis of knowledge of the normative frequency of the test word, but instead with an average approximation of normative frequency that applies to all test words. Thus, if HF words have common features and LF words rare features, the system calculates matching by using an assumption that each test word has features whose diagnosticity lies

[2] The likelihood ratio can be thought of as a measure of the “similarity” between the test cue and the memory trace of each list item. The odds can similarly be thought of as a measure of “familiarity.”

between these two extremes. This assumption is then combined with the default REM assumption that an "old" response is given whenever the calculated "odds" of the test item being old is greater than 1.0. The result is higher HRs and lower FARs for LF words compared with HF words.

One can apply this model to situations like the present one, in which highly similar foils are often used at test. Suppose one ignores the possibility of adjusting the rules for calculating matching, or adjusting the decision criterion for responding "old," on the basis of the knowledge that the test item is an old "type" (and hence either an old target or a similar foil). Then the REM model makes an interesting prediction (explained in detail shortly): The typical mirror effect for normative word frequency should disappear when highly similar foils are used. Basically, LF foils should tend to match the diagnostic rare features in a very similar stored trace and therefore should produce high levels of matching, producing an increase rather than a decrease in the FAR. Hence, REM predicts that the mean probability of a JOF greater than zero, that is, $P(\text{JOF} > 0)$, for LF similar foils will be greater than or equal to the $P(\text{JOF} > 0)$ for HF similar foils, but the typical LF FAR advantage will be observed for dissimilar foils.

The assumptions that give rise to this prediction can of course be questioned (e.g., Norman, 2002; Rotello et al., 2000), so we explore alternative accounts even when this prediction was borne out. In particular we look at augmented REM models that include a recall component, and such models have been proposed to account for the WFE (Balota, Burgess, Cortese, & Adams, 2002; Hirshman et al., 2002; Joordens & Hockley, 2000; Reder et al., 2000). Our data have implications for these models, but we defer that discussion until after we present our findings.

Lastly, we consider the effect of repetitions and word frequency on mean JOFs > 0. As we mentioned earlier, no REM model of JOFs exists. However, Hintzman and Curran (1995) proposed that after an item was judged to be old or new, then a JOF was assigned to the item on the basis of its familiarity. A REM model that makes similar assumptions predicts that mean JOFs > 0 for targets and similar foils will be a nondecreasing function of the number of times an item is presented, and the mean JOFs > 0 will be greater for LF targets than for HF targets. In addition, an interaction similar to the one predicted for false-alarm rates should be observed: The mean JOF > 0 should be greater for HF than LF dissimilar foils, but not for similar foils.

Experiment

In this experiment, singular nouns and an equal number of plural nouns were studied 1, 3, 6, or 12 times. Half of the studied words were HF and half were LF words. After a 30-s addition task, targets, dissimilar foils, and similar foils were presented (similar foils differed from a studied word in their plurality). If the critical features distinguishing targets and similar foils are not learned after 1 or 2 repetitions, then discrimination of targets from similar foils should remain constant for the 3-, 6-, and 12-presentation conditions (roughly speaking, the difference between the HRs and FARs should remain constant). Predictions can be derived for a simple familiarity-based REM model, one assuming that neither criterion nor matching calculations change with degree of familiarity (even when familiarity is so high that there is a virtual certainty that the test item type has been studied). For this model, with reasonable parameter values, the LF HR should be greater than HF HR, the LF FAR for similar foils should be greater than or equal to the HF FAR for similar foils, and the LF FAR for dissimilar foils should be less than the HF FAR for dissimilar foils.

Method

Participants. Seventy-four Indiana University students participated in exchange for course credit.

Design, materials, and procedure. Normative word frequency (HF vs. LF) and study presentations (1, 3, 6, or 12 presentations) were manipulated as within-subject variables. HF words occurred more than 50 times per million and LF words occurred fewer than 10 times and greater than 1 time per million, according to Francis and Kučera (1982), and all words were between 4 and 8 letters in length and could be formed into their plural forms by appending an *s*. Assignment of words to conditions was randomly determined anew for each participant.

For each participant, 32 HF and 32 LF words were randomly chosen to construct the study lists. Each type was randomly subdivided into 16 to be studied in singular form and 16 to be studied in plural form. Each of these groupings was then randomly subdivided into four categories of four words each: words to be presented 1, 3, 6, or 12 times. Each presentation of a given word was in the same plurality. No word was presented twice in a row; 16 untested medium-frequency words ($20 <$ word frequency < 50 per million) were used as filler items in order to help meet this constraint. Each word was presented for 3.0 s in the center of a computer screen and the interstimulus interval was 0.2 s.

Participants were initially instructed that they would view a list of words, and they were to try to remember them for a later unspecified memory test. Plurality was not mentioned at this time. The viewing of the entire study list took approximately 15 min. Following the study list, the participants performed a 30-s math task. Immediately following the math task, the instructions were given for the test portion of the experiment. The test list consisted of 32 targets, 32 similar foils, and 32 dissimilar foils (a word of a given type was tested in either singular or plural form, not both). Target words matched studied words in all respects, including plurality. Similar foils had plurality switched from words studied (for example, if the word *prune* was on the study list, then *prunes* would be the corresponding similar foil). Dissimilar foils did not match studied words (in either plurality). Half the dissimilar foils were presented in each plurality. Assignment of words to all conditions was randomly determined anew for each participant.

At test, the participants were asked to judge how many times a word was presented on the study list. They were specifically warned to pay close attention to the plurality of the words because a difference in plurality between test and study meant a test word was "new" (e.g., they were told that "*cat* is different from *cats*"). They were to indicate their JOF by typing into the computer the number of times they believed it was shown, with the response "0" meaning the word was "new." This portion of the experiment was self-paced.

Results and Discussion

An alpha of .05 was used for all tests. *F* statistics are from repeated measures analyses of variance and *t* tests are two-tailed. The data are shown in Figure 1 and Table 1.

Recognition. Figures 1A and 1B (lines connect the data points) show the hit rates (1A) and the false-alarm rates (1B) plotted as function of the number of times a target or similar foil was presented. Repetitions had a significant effect on both hit rates, $F(1, 73) = 193.3$, $MSE = 0.06$, and false-alarm rates for similar foils, $F(1, 73) = 17.80$, $MSE = 0.09$. That is, the probability of a JOF > 0 increased from 1 presentation to 12 presentations.

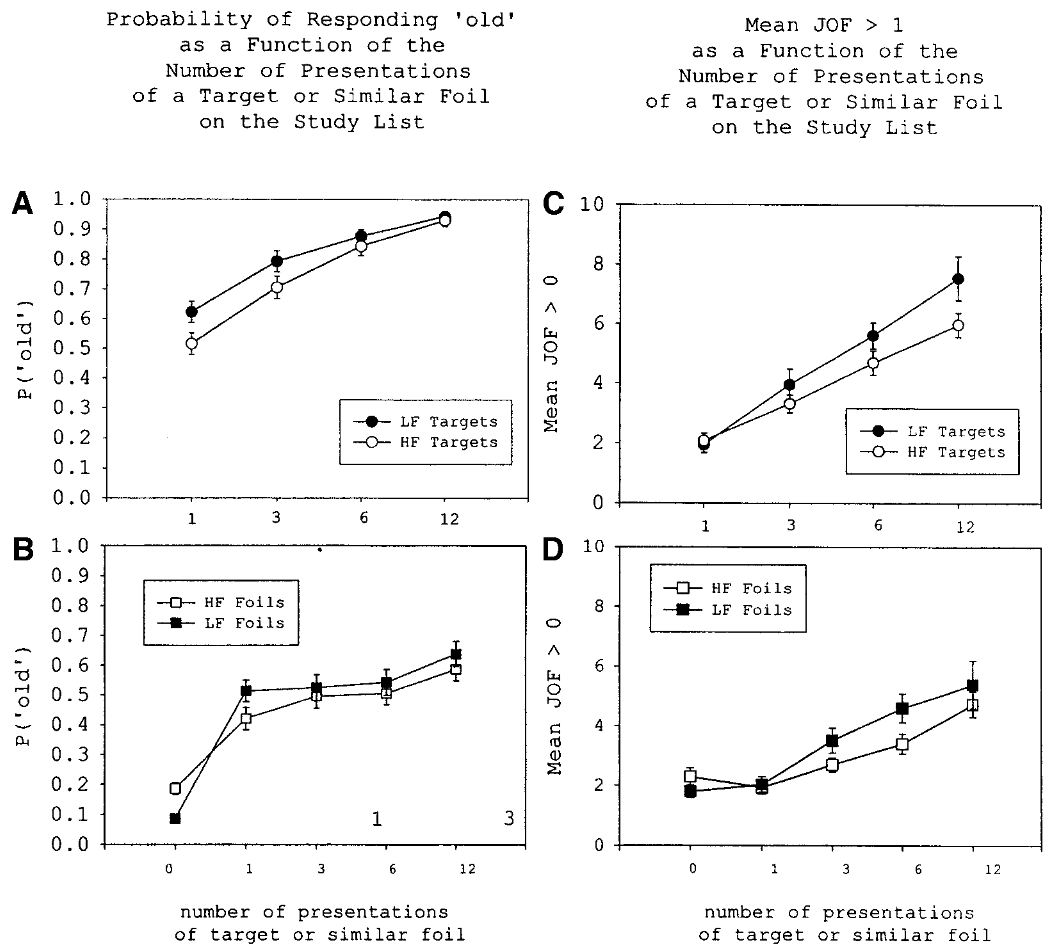

Figure 1. Old–new recognition and judgment-of-frequency (JOF) results of the experiment that manipulated item repetitions, item similarity, and normative word frequency. A: The old–new recognition results in terms of probability of responding "old" (*P*(old)) to targets. B: The old–new recognition results in terms of probability of responding "old" (*P*(old)) to foils. C: The JOF results for targets. D: The JOF results for foils. Error bars represent standard errors. HF = high-frequency words; LF = low-frequency words.

Discrimination of targets from similar foils is measured with d' calculated from HRs and FARs for similar foils.[3] The mean d's are shown in Table 1. Word frequency did not significantly affect d' and did not interact with repetitions (both Fs < 1.0), therefore subsequent statistical analyses are performed on the mean d's collapsed across the word-frequency factor. Overall, d' increased significantly with presentations, $F(1, 73) = 35.42$, $MSE = 2.11$. Planned comparisons indicated that the increases in d' from 1 to 3 presentations and from 3 to 6 presentations were significant, $t(73) = 2.89$ and 2.85, respectively, but not from 6 to 12 presentations, $t(73) = .20$. Although some prior findings suggest that the discrimination of targets from similar foils levels off after just one or two presentations, the point of leveling off in the present study lies somewhere between 4 and 6 presentations. This change may be due to our use of a longer study list, and a consequent lowering of performance.

REM predicts an interaction between repetitions and word frequency such that the typical LF false-alarm rate advantage should

Table 1
Discrimination (d′) of Targets From Similar Foils as a Function of Presentations and Word Frequency

	Word frequency					
	High		Low		*M*	
No. of presentations	*d′*	*SE*	*d′*	*SE*	*d′*	*SE*
1	.39	.19	.36	.18	.37	.15
3	.81	.19	1.00	.20	.91	.14
6	1.35	.16	1.30	.19	1.32	.14
12	1.38	.20	1.21	.20	1.30	.14
M	.98	.11	.97	.12		

[3] Several participants produced hit rates and/or false-alarm rates equal to 1.0 or 0.0. In these cases, we used values of .975 and .025, respectively, to compute *z* scores.

only be observed for dissimilar foils, and this prediction was confirmed, $F(1,73) = 5.79$, $MSE = 0.004$: FARs were greater for HF than LF dissimilar foils, $t(73) = 7.90$, but the FAR for similar foils was significantly greater for LF than for HF words, $F(1, 73) = 4.86$, $MSE = 0.08$. This qualitative pattern of results is consistent with the REM account of the WFE. To fit the results quantitatively, however, we turned to a dual process extension of the REM model (described later).

Mean JOF > 0 results. The mean JOFs > 0 for all participants are shown in Figures 1C and 1D. Mean JOFs > 0 increase in regular fashion with repetitions, for both targets and foils and for HF and LF words. In addition, the mean JOF > 0 is always greater for LF than for HF words, except for the dissimilar foils, for which the opposite is true. These findings are consistent with prior findings showing that JOFs increase with repetitions (Hintzman & Curran, 1995; Hintzman et al., 1992), and they are consistent with the predictions of a familiarity-based REM model for mean JOFs > 0. However, the statistical inferences are limited because some participants did not produce a JOF > 0 in all 18 conditions. Thus, we report the analyses of only complete participant data sets.

The increase in mean JOF > 0 increased significantly for targets and foils, $F(1, 49) = 120.3$, $MSE = 8.17$; $F(1, 15) = 37.6$, $MSE = 7.12$, respectively. The mean JOF > 0 was greater for LF targets than for HF targets but not for foils, $F(1, 49) = 7.03$, $MSE = 5.34$; $F < 1.0$, respectively. The interaction between normative word frequency and target presentations was reliable for targets, $F(1,49) = 10.54$, $MSE = 20.24$. For similar foils the mean JOF > 0 was greater for LF than for HF words, $F(1, 24) = 2.50$, $p \leq .13$, but for dissimilar foils this pattern was reversed, $t(43) = 2.0$, $p \leq .053$. Thus, a mirror-patterned WFE is observed for mean JOFs > 0 only when foils are only randomly similar to studied words. Despite the fact that some effects do not reach significant levels due to relatively small number of observations, these results are consistent with a familiarity-based REM model of mean JOFs > 0.

Hintzman and colleagues (Hintzman and Curran, 1995; Hintzman et al., 1992) observed that the JOFs given to similar foils were distributed in a "bimodal" fashion with tendency to give either a JOF of 0 or a somewhat greater JOF. Figure 2 shows four histograms that plot the number of JOFs of different magnitudes given to similar foils as a function of the number of target repetitions. Our findings are similar to Hintzman and Curran's (1995; Hintzman et al., 1992) in that the distributions of JOFs have two modes: The lower mode occurs at JOF = 0 and the higher mode is greater than 0 and increases with increases in the number of repetitions.

In summary, this experiment produced several key findings: Hit rates rise steadily with presentations. False-alarm rates are low for dissimilar foils and are high for similar foils. False-alarm rates for similar foils rise only slightly as the number of presentations of the similar list item rises from 1 to 12. Hit rates are greater for LF than for HF words. For dissimilar foils, LF words have a lower false-alarm rate, but for similar foils, HF words have a lower false-alarm rate. Mean JOFs > 0 are higher for LF words than for HF words, except for dissimilar foils, for which mean JOFs > 0 are greater for HF words than for LF words.

Modeling the Effects of Repetitions, Similarity, and Normative Word Frequency on Recognition Memory and JOF

In this section, we present two classes of REM models for the present findings. We start by describing the Shiffrin and Steyvers' (1997, 1998) REM model, a model based on a single process of familiarity assessment. We show that the simplest form of this model predicts the present pattern of results in qualitative fashion. It turns out that this model and some one-process variants do not fit the data, and we discuss reasons why we believe this occurs. Then we describe a simple extension of Shiffrin and Steyvers' (1997, 1998) model that incorporates a "dual-process" assumption, and this allows it to predict our findings quantitatively.

REM

The REM representational and encoding processes are assumed regardless of the nature of retrieval. Hence, the following sections on representation and encoding are assumed for both the single- and dual-process models of retrieval that we will discuss shortly.

Representation. REM assumes that generic knowledge about words is stored in separate lexical–semantic memory traces (im-

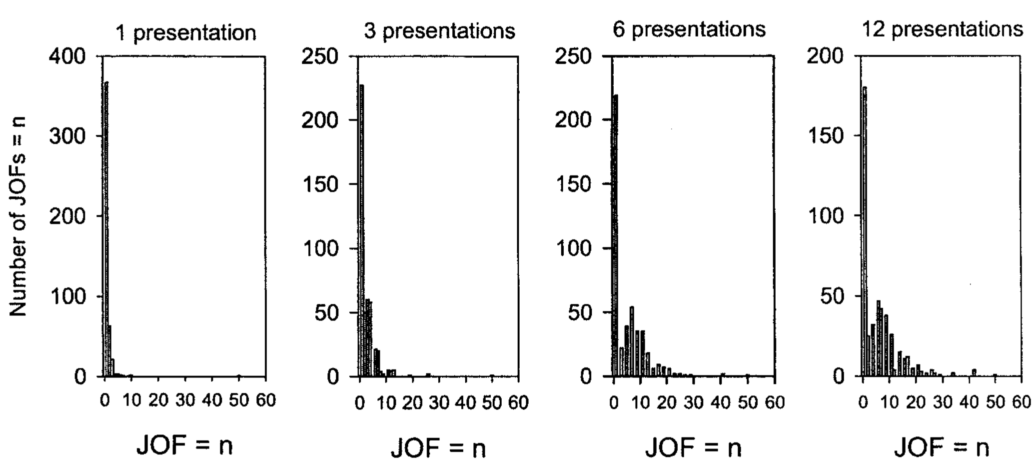

Figure 2. Histograms showing the number of times a judgment of frequency (JOF) equal to *n* was given to similar foils when targets were presented 1, 3, 6, or 12 times.

ages). Each trace consists of a vector, **V,** of w feature values. Each feature value in a lexical–semantic trace is an integer greater than zero; the higher the integer, the lower the base rate probability of that feature in the environment. The base rates for a given feature value, V, are assumed to follow a geometric probability distribution:

$$P(V = j) = (1 - g)^{j-1}g, \qquad j = 1, \ldots \infty \tag{1}$$

The mean feature value and the variability of the feature values are negatively related to g.

Episodic traces. When a word is studied, an incomplete and error prone representation of the word's lexical–semantic trace is stored in a separate episodic image. After t time units of study, the probability that a feature will be stored in the episodic image is $1 - (1 - u^*)^t$, otherwise 0 is stored (u^* is the probability of storing a feature in a unit of time). In our study, a fixed study time was used, so we chose this as the time unit (i.e., $t = 1$), and u^* is the storage probability. A 0 indicates that no feature was stored. If storage of a value occurs, the feature value is correctly copied from the word's lexical–semantic trace with probability c. With probability $1 - c$ the value stored is sampled randomly according to Equation 1.

Repetitions. The rules governing how repetitions affect memory are central to REM (Shiffrin & Steyvers, 1997, 1998). When a word is studied more than once, either (a) a new episodic trace is stored or (b) features are added to an existing trace. Shiffrin and Steyvers (1997) assumed that the second option is more probable than the first in order to account for slightly negative list-strength effects for recognition memory (also see Shiffrin, Ratcliff, & Clark, 1990). When a presentation of a word results in the updating of its previous trace, storage continues as if study time has been added to the earlier presentation; that is, nonzero features replace zero features according to the storage rules described earlier. In the present model, the probability of storing a new trace when an item is repeated is υ (we constrained $\upsilon < .50$, in light of the list-strength findings). Several different traces of a repeated word can of course be stored, and we assume that if information is added to a previous trace, it is added to that trace with the greatest λ_j.[4]

Familiarity-Based Recognition

Old–new recognition. A single-process continuous-state model assumes that recognition is based on the similarity of the test item to contents of memory. This continuous random variable is often referred to as the familiarity of the test item. The REM model described by Shiffrin and Steyvers (1997, 1998) is a single-process familiarity-based model. According to that model, the lexical–semantic vector of w features corresponding to the test item serves as a retrieval cue at test. In the simplest form of the model, the retrieval cue is matched in parallel to the n episodic images (I_j) that were stored during study. For each episodic trace I_j, the system notes which features of I_j match the corresponding feature of the cue, and the value of each feature that matches (n_{ijm} stands for the number of matching values in the jth image that have value i); the system also notes which features mismatch (n_{jq} stands for the number of mismatching values in the jth image). Next, a likelihood ratio, λ_j, is computed for each I_j:

$$\lambda_j = (1 - c)^{n_{jq}} \prod_{i=1}^{\infty} \left[\frac{c + (1 - c)g(1 - g)^{i-1}}{g(1 - g)^{i-1}} \right]^{n_{ijm}}, \tag{2}$$

where λ_j is the likelihood ratio for the jth image and can be thought of as a match-strength between the retrieval cue and I_j. Equation 2 is derived from a ratio of two probabilities: The numerator gives the probability that the retrieval cue and the image represent the same word (in which case features are expected to match, except for errors in storage); the denominator gives the probability that the retrieval cue and the image represent different words (in which case features are expected to match only by chance). The recognition decision is based on the odds, Φ, giving the probability that the test item is old divided by the probability the test item is new (Shiffrin & Steyvers, 1997). This is just the average of the likelihood ratios:

$$\Phi = \frac{1}{n} \sum_{j=1}^{n} \lambda_j, \tag{3}$$

If the odds exceed a criterion, C, then an "old" response is made. The default criterion is 1.0 (which maximizes probability correct), although participants could of course deviate from this setting.

Word frequency. REM predicts a LF HR advantage because the matching of the more uncommon features of LF words produces greater evidence than the matching of the more common features of HF words ("Commonness" is implemented by the assumption that $g_{HF} > g_{LF}$. Malmberg et al., 2002, and Shiffrin & Steyvers, 1997, discuss the REM account of the WFE in greater detail). Similar reasoning leads to the prediction of higher FARs for LF similar foils compared with HF similar foils. For dissimilar foils, however, every feature match is due to chance; such matching occurs more frequently for HF than LF words because HF features are more common (cf. Malmberg & Murnane, 2002). This factor outweighs the higher diagnosticity of matches for the LF words, and HF words are predicted to have higher FARs than LF words (in accord with the data depicted in Figure 1).

Repetitions and similarity. Repetitions allow features to be stored that were not previously stored either by storing features in a new trace or by increasing the number of features stored in an existing trace (Malmberg & Shiffrin, in press; Shiffrin & Steyvers, 1997, 1998). As a result, REM predicts a steady increase in HRs across repetitions. However, the storage of additional features in an existing trace not only strengthens that trace but it also "differentiates" that image from other traces in memory (Shiffrin & Steyvers, 1997, 1998; also see McClelland & Chappell, 1998, and Shiffrin et al., 1990). That is, the trace becomes less similar to most other traces in memory. The idea in REM is simple: For testing of a dissimilar foil, consider the match to a trace of a (different) study item; on average, each extra feature stored in that trace adds more evidence against a match, because the two items being compared are only randomly similar. For dissimilar foils, this factor decreases the FAR across repetitions because memory

[4] On rare occasions, a memory trace corresponding to a word other than the word being studied will produce the greatest activation, and therefore features of the current word will be added to it in error. This simply adds an additional source of noise to episodic memory.

traces and dissimilar retrieval cues tend to mismatch more and more often.

Similar foils are assumed to share a high proportion, σ, of feature values with the matched studied word. Unlike for dissimilar foils, there is one list trace that matches a similar foil in almost all its features. For that one trace, storage of additional features increases evidence of matching because cues with greater similarity values produce greater likelihood ratios. Therefore the FAR for similar foils is greater on average than the FAR for dissimilar foils.[5] It is important to note that only those features that do not overlap may produce differentiation. Hence, increasing target presentations produce greater similar-foil FARs because only a relatively small number of features do not overlap with a target trace.

The exact predictions for the effect of increasing repetitions are therefore dependent on how similar a retrieval cue is to the traces in memory. Figure 3 shows the familiarity-based REM predictions for FAR as a function of the degree of similarity, σ, between a target and a similar foil. (The details of the Monte Carlo simulation and the parameter values for these predictions are presented in Figure 3.) It shows that increasing target presentations increases FARs regardless of the proportion of features shared by a target and a similar foil, and the increase in the FARs is positively related to similarity between the cue and a target trace. For higher similarity values (e.g., .8) the FARs increase sharply with increases in target presentations, and the FARs curves are never truly flat unless the target and foil are only randomly similar or the FARs reach ceiling. Hence, FARs for similar foils are predicted to increase across repetitions because the familiarity produced by the matching features dominates the differentiation produced by mismatching features.

The results from our experiment replicate those of Hintzman and colleagues (Hintzman & Curran, 1995; Hintzman et al., 1992) that showed registration without learning. In addition, the REM model prediction (Shiffrin & Steyvers, 1997) that a mirror-patterned word frequency should occur only when foils are randomly similar to studied items was confirmed. However, the

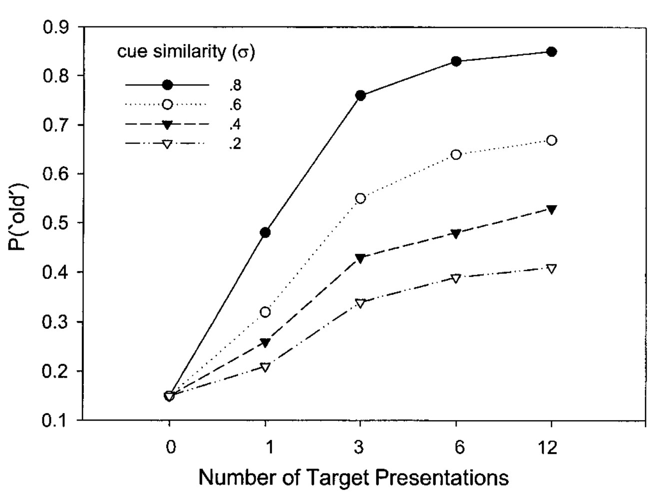

Figure 3. Effect of cue similarity on false-alarm rates, *P*(old), in the single-process familiarity-based retrieving-effectively-from-memory model. Parameter values: $g = .4$; $w = 10$; $t = 1$; $u^* = .5$; $c = .68$; criterion = 1.0; $\upsilon = .05$.

simple familiarity-based REM also predicts a steady increase in FARs for similar foils (see Figure 3), and this is inconsistent with our findings (see Figure 1B). However, a relatively steady increase in JOFs with increases in presentations was observed in the Experiment (see Figures 1C and 1D), and hence the rise in similar-foil JOFs with increases in target presentations can be accounted for by assuming that JOFs are based on a monotonic transformation of an item's familiarity. Below, we specify one way that this might be accomplished. Before doing so, however, we first briefly explore one straightforward way the familiarity-only model might be modified in order to account for registration-without-learning, and show that even with the added complexity, it cannot predict all the data shown in Figure 1.

One obvious way the familiarity-based REM model might be augmented to predict the flattening FARs and the steady increase in JOFs for similar items is to assume that differentiation plays a relatively larger role for old–new recognition than it does for JOFs. That is, the differentiation mechanism might be exploited for old–new recognition in order to counteract the increase in familiarity produced by increasing item repetitions. For instance, the proportion of old responses (i.e., JOFs > 0) might reflect a heightened level of attention paid to the salient differences (in the present case the test item's plurality) between targets and similar foils, and the mean JOFs given to an item called old might reflect its overall familiarity (cf. Figure 2).

Specifically, assume that memory is probed twice when a test item is presented, producing two different levels of familiarity associated with two different retrieval cues. Retrieval cues for target items will match the contents of at least one trace in memory except when encoding failed to store (or accurately store) a feature. Retrieval cues for new items are only randomly similar to the contents of memory, but retrieval cues for similar foils consist of features that could potentially match a target item's trace or traces and a smaller proportion of features that will mismatch the features in a target trace (except as the result of storage errors). A first probe with only the non-overlapping features of the test item (in this case non-plurality features) produces an initial level of familiarity, which if it does not exceed a subjective criterion, leads the test item to be rejected. If, however, the familiarity of the overlapping features exceeds the criterion, then a second probe is made with a retrieval cue containing both the overlapping and the potentially non-overlapping features. Weighting more heavily in the decision process the evidence that is associated with the non-overlapping features serves to focus attention on the salient potential difference between targets and similar foils. By more

[5] It is important to note that the calculation of likelihood ratio is the same for all items (i.e., on the basis of the assumption that foils are all dissimilar). One could imagine alternative systems for experiments like the present one. For example, new calculations could be derived under the assumption that foils have a probability p of being similar and a probability $1 - p$ of being dissimilar. Alternatively, one might imagine a system in which an initial calculation of familiarity is carried out; if a very high odds is found, then this could be used as an indicator the test word had been studied in either its singular or plural form, and a new calculation could be carried out on the assumption that the only possibilities are target and similar foil. Such systems would considerably change the predictions given above for HRs and FARs.

heavily weighting the evidence produced by the potentially mismatching features, the differentiation process might be facilitated.

We implemented these assumptions in the single-process REM model by assuming that an odds associated with the overlapping features, $\Phi_{S,}$ is initially computed and compared with a criterion that must be exceeded or else the item is rejected. That is, the initial memory probe involves matching against the contents of memory only the features that overlap between targets and similar foils. In the present case, these are the features not used to represent a word's plurality. The plurality features are not used to probe memory at this time because they alone do not provide any evidence that an item is old or new (because equal numbers of singular and plural words were studied and tested). If Φ_S exceeds the criterion, then memory is probed a second time with a retrieval cue containing both the shared features and the potentially unshared features such that odds associated with the shared features, Φ_S, is weighted less than the odds for the potentially mismatching features, Φ_0. Both the overlapping and non-overlapping features are used in this probe to determine whether a particular word was studied in its singular or plural form. The odds associated with the non-overlapping features were weighted, a, in the following manner to produce an overall level of familiarity: $\Phi = \Phi_s \cdot \Phi_0^{1+a}$, where $0 \leq a$.

The critical question for the modified familiarity-based model is whether weighting the matching evidence associated with the overlapping features can enhance the differentiation process enough to produce increasingly flat FARs as a function of the number of target presentations. Our modeling efforts found that this is not the case: The FARs are predicted to increase with increases in presentations even though targets and similar foils were assumed to overlap in only half of their features, which gives the present model a fighting chance, but also seems somewhat inconsistent with the spirit of the operational definition of "highly similar." Other weighting functions were explored, but they produced results similar increases in FARs with increases in target presentations.[6]

In summary, both of the familiarity-based models that we described can in principle handle the pattern of JOFs observed in our experiment (see also Hintzman & Curran, 1995; Hintzman et al., 1992), but neither of the familiarity-based models of recognition that we considered predict the flat FARs that we observed. For now, we leave that JOF calculation unspecified, deferring the specifics while we address how adding a recall process to the single-process REM model can help account for the recognition findings.

As discussed earlier, a recall-like process may contribute significantly to old–new recognition performance in tasks like the present one (e.g., Curran, 2000; Hintzman et al., 1992; T. C. Jones & Jacoby, 2001; Kelley & Wixted, 2001; Rotello et al., 2000). It may be, for example, that a critical feature like plurality does not by itself change familiarity enough to allow sufficient discrimination of targets from similar foils. However, recall of a particular episodic trace might allow the contents of that trace to be examined and plurality explicitly assessed. Note that such a recall process could operate to counteract the expected effects of familiarity (Jacoby, 1991): As repetitions increase, the trace or traces become stronger (on average), causing familiarity of the similar foil to rise but also causing recall to improve to provide a better assessment of plurality, allowing correct rejection of the otherwise familiar similar foils. We next give a quantitative implementation of a simple version of this approach within the REM framework.

Combining familiarity and recall. The dual-process REM model we propose is similar in concept to those listed above. We start with the basic REM model, and use it to calculate an odds value, interpreted as familiarity. If familiarity is less than a criterion, C, then a "new" response is given. Otherwise, an attempt is made to recall the feature value that encodes plurality.

Recall in REM (Diller, Nobel, & Shiffrin, 2001; Malmberg & Shiffrin, in press; Shiffrin & Steyvers, 1998) is conceptualized to occur in much the same way it occurs in SAM (Raaijmakers & Shiffrin, 1980, 1981), as a search process involving cycles of sampling, recovery, and decision. The general recall model involves a quite complex interaction of long-term memory, short-term memory, and control processes (Atkinson & Shiffrin, 1968; Raaijmakers & Shiffrin, 1980, 1981). For our present purposes, however, we can work with a greatly simplified model. The recall part of the present dual-process REM model is based on the following background assumptions (not modeled explicitly): Retrieval cues are used to probe memory; traces are stochastically sampled with replacement; and the more similar a trace is to the retrieval cue (i.e. the likelihood ratio), the more likely it will be sampled and the less likely other traces will be sampled. When an image is sampled, an attempt is made to recover its contents. The probability of recovering the contents is positively related to the number of correctly stored features. If the contents are recovered, then a response is formulated. If the contents are not recovered and a criterion number of attempts have not been made, the sampling and recovery process begins anew.

Old–new recognition. For present purposes, all this background reasoning can be boiled down to the following explicit assumption: For each item studied r times, there is a probability, q, that the test item's image is sampled and the critical feature is recalled.

$$\hat{q} = ap^{1/r}[(c(1 - (1 - u^*)^r)], \tag{4}$$

where p and a determine an overall level of recall success and its utilization; c is the probability of correctly storing a feature; u^* is probability of storing a feature on a given attempt to do so[7]; and r equals the number of times a target was presented for similar foils (the target item is recalled). The first part of Equation 4 reflects an increasing probability of recovering sampled features with increases in r, and the second part says that the probability of sampling the correct image will increase with r to a level asymptoting at the probability that a feature will be copied correctly. When recall fails, we assume that the participant guesses "old" with probability γ and then generates a JOF > 1 on the basis of the item's familiarity.

[6] In fact, weighting the non-overlapping features actually increases the FARs because accidentally matching one of these features due to errors in storage or due to a match from a different trace or traces adds more to the level of familiarity than mismatching a feature takes away according to Equation 2.

[7] We refer to this value as $\hat{q}$ because this value slightly underestimates the probability of recalling the critical features because in principle one might recall features that were stored correctly only by chance and this can lead to correctly recalling critical features. However, this is very unlikely, and it greatly simplifies matters to work with $\hat{q}$.

JOFs. The pattern of JOFs that we observed confirmed in a qualitative manner the pattern of odds (or familiarity) generated by simple single-process REM model of repetition and normative WFEs (Malmberg et al., 2002; Shiffrin & Steyvers, 1997, 1998). To generate JOF quantitative predictions, we assume that the mechanism for generating JOFs is similar to that for generating confidence ratings (Shiffrin & Steyvers, 1997): An ordered set of criteria is maintained that reflects different levels of familiarity, and JOFs are made by comparing the criteria to the odds associated with the test item: The JOF given is that which is associated with the greatest criterion exceeded by the familiarity of the test item. It is possible that several types of functions might explicitly relate the distributions of odds to different JOF values depending on how they are parameterized, but for present purposes we chose the following simple one:

$$C_i = b(1 - e^{-ri}), \tag{5}$$

where C_i is the log odds associated with JOF $= i$, b is a scaling parameter, and r is a rate parameter. The JOF given is that integer value, i, associated with the criterion value of the greatest criterion exceeded by the item familiarity. For present purposes we assumed that 24 criteria were used (cf. Figure 2).[8]

Figures 4 and 5 show a fit of the dual-process REM model to the recognition and JOF data, respectively.[9] They indicate that the dual-process REM model captures the data reasonably well and accounts for all of our major findings: The HRs increase because repetitions cause storage of more target traces and more feature values in the target trace or traces; both factors increase the likelihood ratios for targets and hence the tendency to exceed the old–new criterion on the basis of their familiarity and the probability that the recall process will be invoked in order to check the targets' pluralities. According to Equation 4, increases in repetitions of targets that exceed the old–new criterion increase the chance that the critical features that match the test item will be recalled, thus resulting in increases in the number of JOFs > 0 or "old" responses. Because the familiarity of the targets increases with increases in repetitions, so too should the mean JOF > 0 given to targets according to Equation 5.

Figure 4B shows that the FARs for similar foils is greater than the FAR for new foils because similar foils tend to produce greater levels of familiarity and therefore are more likely to exceed the old–new criterion. The predicted FAR for a similar foil does not, however, increase steadily with repetitions of its corresponding studied word. Repetitions cause the similar foil to become more familiar but also cause the recall process increasingly to detect a mismatch of plurality, according to Equation 4. The second factor counteracts the first: Repetitions cause an increase in the probability that a similar foil exceeds the *yes–no* criterion, but they also cause an increase in the probability of responding "new" because of recall of a plurality mismatch. Nonetheless, the monotonic increase in the familiarity of the similar foils should produce a prediction of a steady increase in the mean JOF > 0, and this is demonstrated in Figure 5B.

The interactions of these predictions with normative word frequency have been discussed already: The HRs are greater for LF than for HF words because they tend to consist of more uncommon or diagnostic features, which in turn produce greater evidence that item is old, and the FARs are greater for HF than for LF words because common features tend to accidentally match more features

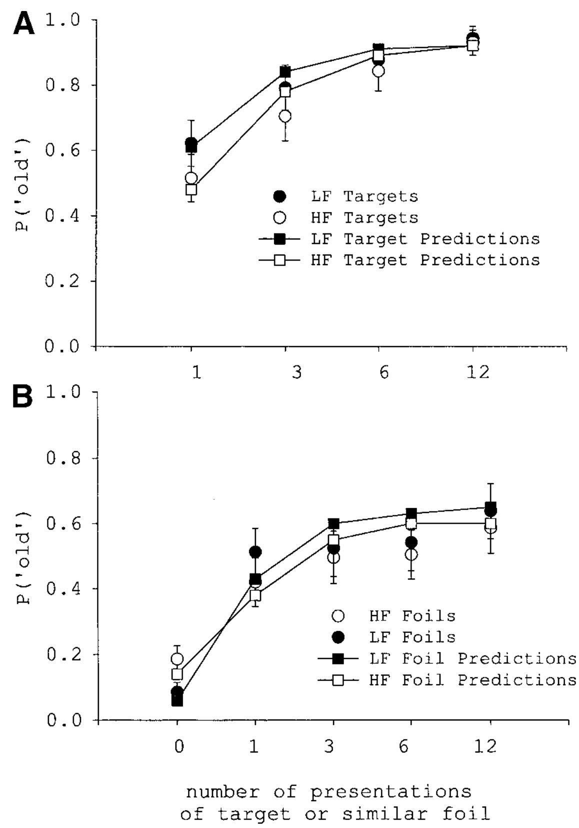

Figure 4. Probability of responding "old" as a function of the number of presentations of a target or similar foil in the study list: Dual-process retrieving-effectively-from-memory model predictions for "old–new" recognition. A: Hit rates. B: False-alarm rates. Error bars are 95% confidence intervals. Predictions are based on a Monte Carlo simulation of 500 participants with the following parameter values: g = .375, g_{HF} = .40, g_{LF} = .375, w = 10, u^* = .48, c = .77, υ = .20, γ = .90, C = 1.6, p = .50, a = .37. HF = high-frequency words; LF = low-frequency words.

in memory than diagnostic features. The reversal of the frequency effect for similar foils (compared with dissimilar foils) is due to the fact that the highly diagnostic LF features tend to match the features in the corresponding trace, and thereby increase the likelihood that the test item is a target. The predictions closely mimic the data, as shown in Figure 4. According to the present models,

[8] It seems unlikely that every participant sets the same number of criteria. However, because we are not fitting individual participants' data, we assume that they do for simplicity. It might be that the number of criteria set reflects the capacity of the participant to maintain a certain number of criteria (e.g., 7 ± 2). In fact, the distribution reported by Hintzman and Curran (1995) might suggest that a smaller number of criteria are used, and that relatively high criteria correspond to JOFs that increase in increments greater than 1. Our distributions (shown in Figure 2) provide less compelling evidence for such a model, and therefore we choose to work with simpler function relating familiarity to JOFs in increments of 1.

[9] This is not necessarily the best fit of the model to the data. The single-process models can be rejected on the basis of their inability to qualitatively fit the recognition data, as discussed in the text.

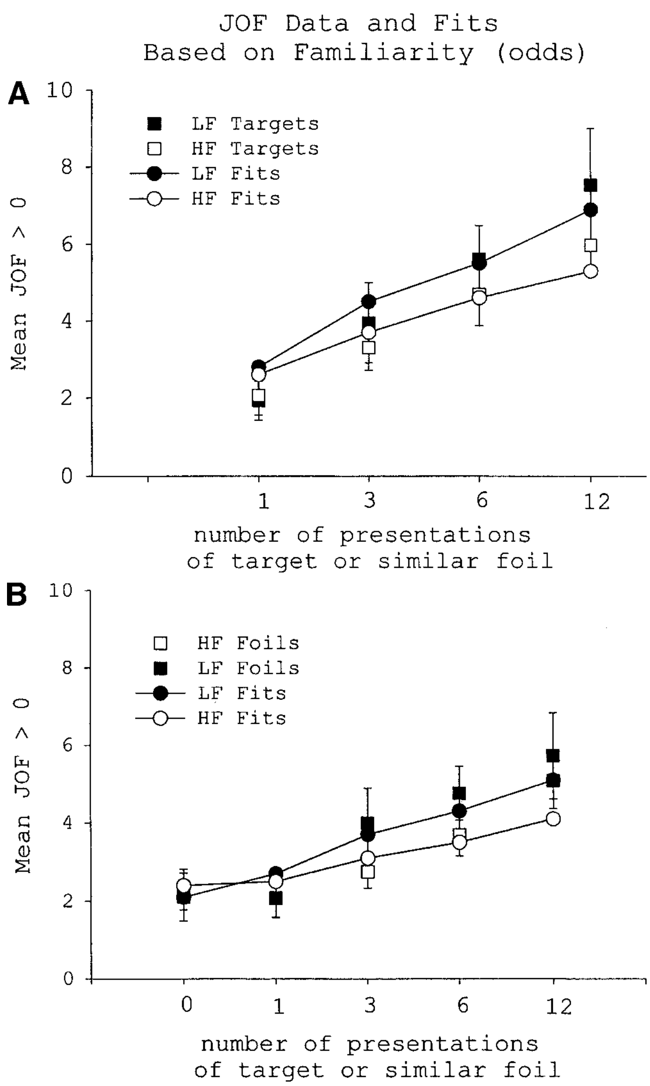

Figure 5. Retrieving-effectively-from-memory model predictions for judgments of frequency (JOFs). A: Hit rates. B: False-alarm rates. Error bars are 95% confidence intervals. Predictions are based on a Monte Carlo simulation of 500 participants with the following parameter values: g = .375, g_{HF} = .40, g_{LF} = .375, w = 10, u^* = .48, c = .77, υ = .20, γ = .90, C = 1.6, p = .50, a = .37, b = 15, r = .20. HF = high-frequency words; LF = low-frequency words.

the mean JOFs > 0 should conform to the level of familiarity associated with a given item. Specifically, mean JOFs > 0 should increase with repetitions, the mean JOF > 0 should be greater for LF targets and LF similar foils than for HF targets and HF similar foils, but the mean JOF > 0 should be less for LF than for HF new foils. Again, Figure 5 shows that the dual-process REM model's predictions provide a reasonable account of the JOF data.

General Discussion

This research project had two main objectives: One objective was to observe the interaction between word similarity and normative word frequency. The REM model predicts a mirror-patterned WFE only when foils are randomly similar to studied items. When foils are similar to at least one studied item, the usual LF FAR advantage observed for randomly similar foils should be reversed. This prediction, made by both the single-process REM model of Shiffrin and Steyvers (1997) and the dual-process model that we presented in this article, was confirmed. To make this prediction, the critical factor is the assumption that LF words tend to consist of more diagnostic features than HF words, an assumption that has been confirmed by prior research (Malmberg et al., 2002).

It is interesting to note, however, that for the dual-process model we did not find it necessary to assume that the recall process produces an advantage for HF words. This alternative assumption could have been considered because HF words are better freely recalled than LF words from pure lists (Gillund & Shiffrin, 1984), but this assumption is not appealing because there is little or no free recall advantage for HF words when mixed lists are used (Gillund & Shiffrin, 1984), as in the present study. Additionally, the assumption of an HF recall advantage in our simple model predicts that the discrimination of targets and similar foils should be greater for HF than for LF words, but Table 1 shows that this is not the case. Perhaps the best reason for assuming that the recall process is not sensitive to normative word frequency is that the plurality is logically independent of normative frequency: Other features of LF words may be rarer and more diagnostic, but this does not seem possible for the plurality feature.

Within the framework of REM, the finding that the typical LF FAR advantage can be reversed when foils are similar to at least one target item is straightforward, but this might not be the case for models that predict mirror effects on the basis of a "concentering" of the theoretical distributions that underlie performance (e.g., Glanzer, Adam, Iverson, & Kim, 1993). In these models, a variable that increases FARs also decreases HRs. Our data show, however, that only the FAR effect, and not the HR effect, is reversed by repetitions (also see Malmberg & Murnane, 2002). Perhaps concentering models are amenable to independent shifts in the target and foil distributions if they also added a dual-process assumption, but further analysis of this possibility is left for future research.

"Context-noise" models might also have a difficult time explaining the present results. According to Dennis and Humphries's (2001) bind, cue, and decide memory theory (BCDMEM), each word has a separate lexical–semantic trace that is associated with the different contexts in which it has been encountered, and recognition is based on the similarity between these contexts and the context in which memory is tested. The LF FAR advantage is observed because LF words tend to have appeared in relatively few preexperimental contexts. Hence, it is relatively easier to reject a LF than a HF foil because the contexts in which LF foils have been encountered are less similar to the test context than the contexts in which HF foils have been found. If singular and plural forms of a word have the same trace, it is not apparent why the discrimination of targets from similar foils is greater than chance in the present experiment. One might argue within the framework of BCDMEM that singular and plural forms of a word have different lexical–semantic traces. However, such a model would predict that discrimination based on plurality would be good, whereas our data show relatively poor discrimination even after 12 presentations of the target. Thus, although the present findings are consistent with the REM model's predictions, it is not clear how concentering and context-noise models can predict our recognition findings.

It is worth noting that the current REM account of the WFE is a single-process account; the mirror pattern is the result of an interaction of the global-matching mechanism and certain representational assumptions (discussed at length above). The recall mechanism does not favor LF words.[10] Other dual-process models posit that LF words are more recallable than HF words, and this produces the LF HR advantage (Joordens & Hockley, 2000; Reder et al., 2000). The LF FAR advantage is a result of the LF words being inherently less familiar than HF words due to the frequency or recency with which they have been encountered prior to the experiment. Although most dual-process models of the WFE are not specified in a form that allows for concrete predictions, the Source of Activation Confusion (SAC; Reder et al. 2000) model assumes that memory consists of word nodes, context nodes, and links between them. The HF word nodes have a higher base-rate activation level than LF words, and hence, the result is a greater level of activation when memory is probed with a HF foil. When a word event is encoded an associative link is created between the appropriate word node and a context node. Because HF words have been encountered more often, there exist more word-to-context links for HF words than for LF words. Probing memory with a LF word leads to a greater chance of recalling the event because activation is limited and hence a greater concentration of activation spreads from the LF word node to the experimental context node.

Although the model we described accounts for the present finding that the LF FAR advantage reverses when foils are highly similar to targets, this finding is complicated to explain within the extant SAC framework. Critically, SAC must make assumptions about whether singular and plural versions of a word are represented by a single word node or by two different word nodes. A simple single-node model would predict that targets and similar foils cannot be discriminated, and our findings rule out this possibility. A strong version of a two-node model would assume that the plurality of test item determines whether the encoding strengthens the baseline activation of the singular or plural node. If only the node corresponding to the plurality of the studied word is strengthened, however, then the FAR for similar foils should be the same as the FAR for dissimilar foils (because it would not be strengthened when the target is encoded), a prediction which is disconfirmed by the present findings.

Hence, our findings suggest within the SAC framework that encoding a word involves strengthening both the baseline activations of singular and plural word nodes. Strengthening the baseline activation of the opposite form of the word accounts for higher FARs for similar than for dissimilar foils and the increase in JOFs made to similar foils with increases in target repetitions. A complication arises, however, because presumably a word-to-context associative link is created for both forms of the word at the same time the baseline activations of both the singular and plural word nodes are strengthened. If so, it is unclear why the FAR for similar foils would level off after 2 or 3 presentations because the recall process would not be able to effectively reject a word on the basis of recollection of its plurality.

The Effect of Target–Foil Similarity on Recognition Memory

The second objective of this research was to develop a formal account within the REM framework for the registration-without-learning phenomena discovered by Hintzman and colleagues (Hintzman & Curran, 1995; Hintzman et al., 1992). Specifically, we sought to determine whether the differentiation mechanism inherent to REM could produce a FAR function for similar foils that did not rise steadily with increasing numbers of target presentations, whether a model that allowed for additional storage after the first 2 or 3 presentations could explain why discrimination of targets from similar foils did not continuously improve increases in target presentations, and whether an item's familiarity could be basis for a JOF in paradigm like one discussed here.

Like McClelland and Chappell (1998), we initially thought that a single process familiarity model could handle results like those of Hintzman and Curran (1995) and those in the present article. We suspected increases in differentiation with increases in repetitions might allow for a negatively accelerating or an inverted U-shaped FAR function. However, these intuitions were disconfirmed. For the REM models that we have investigated, we found that the increase in activation produced by matching the features that are not associated with an item's plurality outweighed the decrease in activation caused by mismatching the features that do represent an item's plurality (see Figure 2). Similar conclusions were made by McClelland and Chappell (1998), for their model. For other paradigms, like the mixed–pure lists paradigm used to explore list-strength effects (Murnane & Shiffrin, 1991; Norman, 2002; Ratcliff, Clark, & Shiffrin, 1990), differentiation is effective because targets and foils are only randomly similar. When targets and foils share features, however, the activations associated with similar foils increase with increases in target presentations. Hence, a familiarity-based account of the recognition could not explain why there was little or no increase in FARs after a target had been presented more than once.

Combined with the findings that show a steady increase in JOFs with increases in target presentation, this led us to a simple dual-process augmentation of the REM model in order to predict the registration-without-learning findings. The dual-process REM model that we described assumes that a recall mechanism is invoked if the familiarity of an item exceeds a subjective criterion and if the recalled features match those of the test item, an "old" response or a JOF > 0 is given. If an "old" response is made, the JOF given is a strictly increasing function of the item's familiarity.

The Registration-With-Learning Hypothesis

Our modeling shows that the results from this experiment are consistent with the assumption that repetitions increase the storage of features that discriminate between targets and similar foils. On

[10] A well-known finding is that free recall is better for HF than for LF words when a list consisting of both HF and LF words is studied. According to the SAM model (Gillund & Shiffrin, 1984), there are more associative retrieval paths for HF words on a list, which produces a greater chance for sampling and recovering HF word events. Use of such retrieval strategies for yes-no recognition seems inappropriate because the to-be-remembered item is in fact already given.

the other hand, in many situations there are limits in REM on the benefits of repetitions. When repetitions accumulate in an existing trace, previously stored features are ordinarily left alone, even if stored incorrectly. Thus feature storage in these situations is limited by the probability of storing features and storing them correctly (cf. Equation 4). This limit would apply to the feature (or features) that encodes plurality. However, REM allows incorrectly stored features to be corrected if attention is directed to the feature in question, in which case the discrepancy would be noted and a chance at correction would occur (Shiffrin & Steyvers, 1997). This correction mechanism is consistent with Hintzman and Curran's (1995) finding that additional learning can be induced on later repetitions when participants were given explicit feedback concerning their knowledge of plurality.

Conclusions

The theoretical proposition that a recall-like process occurs during recognition testing is quite old (e.g., Atkinson & Juola, 1974; Mandler, 1980; Yonelinas, 1994; see Jacoby, 2000; Hintzman, 2000; Yonelinas, 2003, for reviews) and certainly fits with the subjective impressions of anyone participating in a recognition memory study. Gillund and Shiffrin (1984) showed that a model based only on familiarity could handle much of the data existing at that time, and hence the dual-process account that we propose to account for the present findings might not be necessary to account for other recognition memory phenomena (cf. Malmberg, 2002; Malmberg et al., 2004). In our view, the recall mechanism in the present model is useful because it provides information that leads to a different decision than the information provided by the familiarity process. Specifically, the recall mechanism provides a means for discriminating targets from similar foils, even though both types of items are highly familiar. More generally, we speculate that recognition performance will be increasingly influenced by the outcome of a recall-like process to the extent that targets and foils are similar and the features that distinguish them are known. In our case, we believe this salient aspect is the plurality of the items (cf. Rotello, et al., 2000). In other cases, this could be the modality of presentation (e.g., Hintzman & Caulton, 1997; McElree, Dolan, & Jacoby, 1999), voice of the presenter (e.g., Hintzman, Caulton, & Levitin, 1998), a paired associate (e.g., Diller et. al., 2001; T. C. Jones & Jacoby, 2001), and/or the temporal position between different lists (e.g., Hintzman, Caulton, & Levitin, 1998; Jacoby, 1991). What many of these experiments have in common with our experiment is that some foils are known to be similar to targets and there exists a salient aspect to the study events that if remembered at test provides direct evidence that the test item either appeared or did not appear in the experimenter-specified context. As a package, these experiments have produced empirical findings that are difficult to handle parsimoniously with single-process models but that fit quite well with appropriately specified dual-process models.

References

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology: Vol. 1. Learning, memory, and thinking* (pp. 243–293). San Francisco: Freeman.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In W. E. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89–195). New York: Academic Press.

Balota, D. A., Burgess, G. C., Cortese, M. J., & Adams, D. R. (2002). Memory for the infrequent in young, old, and early stage Alzheimer's disease: Evidence for two processes in episodic recognition performance. *Journal of Memory and Language, 46,* 199–226.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review,* 3, 37–60.

Crowder, R. G. (1976). *Principles of learning and memory.* Hillsdale, NJ: Erlbaum.

Curran, T. (2000). Brain potential of recollection and familiarity. *Memory & Cognition, 20,* 923–938.

Curran, T., & Cleary, A. (2003). Using ERPs to dissociate recollection from familiarity in picture recognition. *Cognitive Brain Research, 15,* 191–205.

Dennis, S., & Humphries, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108,* 452–478.

Diller, D. E., Nobel, P. A., & Shiffrin, R. M. (2001). An ARC–REM model for accuracy and response time in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 414–435.

Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar.* Boston: Houghton Mifflin.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 12,* 8–20.

Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review, 100,* 546–567.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace model. *Psychological Review, 95,* 528–551.

Hintzman, D. L. (2000). Memory judgments. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 165–177). New York: Oxford University Press.

Hintzman, D. L., & Caulton, D. A. (1997). Recognition memory and modality of judgments: A comparison of retrieval dynamics. *Journal of Memory and Language, 37,* 1–23.

Hintzman, D. L., Caulton, D. A., & Levitin, D. J. (1998). Retrieval dynamics in recognition and list discrimination: Further evidence of separate processes of familiarity and recall. *Memory & Cognition,* 26, 449–462.

Hintzman, D. L., & Curran, T. (1995). When encoding fails: Instructions, feedback, and registration without learning. *Memory & Cognition, 23,* 213–226.

Hintzman, D. L., Curran, T., & Oppy B. (1992). Effects of similarity and repetition on memory: Registration without learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 667–680.

Hirshman, E., Fisher, J., Henthorn, T., Arndt, J., & Passannante, A. (2002). Midazolam amnesia and dual-process models of the word-frequency mirror effect. *Journal of Memory and Language, 47,* 499–516.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513–541.

Jacoby, L. L. (2000). Recollection and familiarity. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 215–228). New York: Oxford University Press.

Jones, C. M., & Heit, E. (1993). An evaluation of the total similarity principle: Effects of similarity on frequency judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 799–812.

Jones, T. C., & Jacoby, L. L. (2001). Feature and conjunction errors in recognition memory: Evidence for dual-process theory. *Journal of Memory and Language, 45,* 82–102.

Joordens, S., & Hockley, W. E. (2000). Recollection and familiarity through the looking glass: When old does not mirror new. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1534–1555.

Kelley, R., & Wixted, J. T. (2001). On the nature of associative information in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 701–722.

Malmberg, K. J. (2002). On the form of ROCs constructed from confidence ratings. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 380–387.

Malmberg, K. J., & Murnane, K. (2002). List composition and the word-frequency effect for recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 616–630.

Malmberg, K. J., & Shiffrin R. M. (in press). The one-shot hypothesis and the list-strength effect for free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Malmberg, K. J., Steyvers, M., Stephens, J. D., & Shiffrin, R. M. (2002). Feature frequency effects in recognition memory. *Memory & Cognition, 30,* 607–613.

Malmberg, K. J., Zeelenberg, R., & Shiffrin, R. M. (2004). Turning up the noise or turning down the volume? On the nature of the impairment of episodic recognition memory by Midazolam. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 537–546.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87,* 252–271.

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory, *Psychological Review, 105,* 724–760.

McElree, B., Dolan, P. O., & Jacoby, L. L. (1999). Isolating the contributions of familiarity and source information to item recognition: A time-course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 563–582.

Murnane, K., & Shiffrin, R. M. (1991). Interference and the representation of events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 855–874.

Norman, K. A. (2002). Differential effects of list strength on recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 1083–1094.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 14, pp. 207–262). New York: Academic Press.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). The list-strength-effect I: Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 163–178.

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember–know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 294–320.

Rotello, C. M., MacMillan, N. A., & Van Tessel, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory and Language, 43,* 67–88.

Rugg, M. D., Cox, C. J. C., Doyle, M. C., & Wells, T. (1998). Event-related potentials and the recollection of low and high frequency words. *Neuropsychologia, 33,* 471–484.

Sheffert, S. M., & Shiffrin, R. M. (2003). Auditory registration without learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 10–21.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6,* 156–163.

Shiffrin, R. M., Huber. D. E., & Marinelli, K. (1995). Effects of category length and strength on familiarity in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 267–287.

Shiffrin, R. M., Ratcliff, R., & Clark, S. E. (1990). List-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 179–195.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM retrieving effectively from memory. *Psychonomic Bulletin & Review, 4,* 145–166.

Shiffrin, R. M., & Steyvers, M. (1998). The effectiveness of retrieval from memory. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 73–95). London: Oxford University Press.

Wixted, J. T. (1992). Subjective memorability and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 681–690.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1341–1354.

Yonelinas, A. P. (2003). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46,* 441–517.

Received March 15, 2002
Revision received September 3, 2003
Accepted September 18, 2003 ■