Journal of Experimental Psychology:
Learning, Memory, and Cognition
2005, Vol. 31, No. 3, 568–578

0278-7393/05/$12.00 DOI: 10.1037/0278-7393.31.3.568

OBSERVATION

The Effects of Word Frequency and Similarity on Recognition Judgments: The Role of Recollection

Heekyeong Park, Lynne M. Reder, and Daniel Dickison
Carnegie Mellon University

K. J. Malmberg, J. Holden, and R. M. Shiffrin (2004) reported more false alarms for low- than high-frequency words when the foils were similar to the targets. According to the source of activation confusion (SAC) model of memory, that pattern is based on recollection of an underspecified episodic trace rather than the error-prone familiarity process. The authors tested the SAC account by varying whether participants were warned about the nature of similar foils and whether the recognition test required the discrimination. More false alarms for low-frequency similar items occurred only when participants were not warned at study about the subtle features to be discriminated later. The differential false-alarm rate by word frequency corresponded to the pattern of remember responses obtained when the test instructions did not ask for a subtle discrimination, supporting the SAC account that reversed false-alarm rates to similar foils are based on the recollection process.

Keywords: word frequency, item similarity, recognition, recollection, encoding instructions

When words with different normative frequencies are studied, low-frequency words are better recognized than high-frequency words, and low-frequency words show lower false-alarm rates than high-frequency words. This effect has been termed the *word frequency mirror effect* (Glanzer & Adams, 1985). The word frequency mirror effect has been extensively studied in recognition (e.g., Arndt & Reder, 2002; Balota, Burgess, Cortese, & Adams, 2002; Glanzer & Adams, 1990; Guttentag & Carroll, 1997; Hintzman, 1994; Hirshman, Fisher, Henthorn, Arndt, & Passanante, 2002; Joordens & Hockley, 2000; Malmberg, Steyvers, Stephens, & Shiffrin, 2002; Murdock, 1998, 2003; Reder, Angstadt, Cary, Erickson, & Ayers, 2002), and many theoretical accounts have been proposed for it. Reder and her colleagues (Reder et al., 2000) suggested a mechanism for the word frequency mirror effect based on a dual-process model of recognition called *source of activation confusion* (SAC).

According to the SAC account for the word frequency mirror effect, the words of different normative frequencies have different numbers of prior contextual associations that fan out from their respective concept nodes. Differential fan affects the amount of activation that spreads from the concept node to any specific memory trace such that the more links that share activation, the less activation reaches any one node. Low-frequency words have a smaller contextual fan because they have been seen less frequently. Therefore, the activation that spreads to the relevant study episode is greater when the source node sending activation is low frequency (fewer competing links) rather than high frequency (more competing links). Because less activation spreads to the relevant episode node from high-frequency words, their associated episode nodes are less likely to pass threshold for a recollection. As a consequence, judgments for high-frequency words will more often be based on the familiarity of the concept itself. Support for this account of the word frequency mirror effect came from the remember–know paradigm such that recognition of low-frequency words was expressed with more remember responses, whereas recognition of high-frequency words was expressed with more know responses (Reder et al., 2000). In addition, SAC also predicts more false alarms for high-frequency foils than for low-frequency foils as manifested by know judgments. This is because high-frequency words are more familiar than low-frequency words.

Recently, Malmberg, Holden, and Shiffrin (2004) showed that the typical advantage of lower false-alarm rates for low-frequency words was reversed when similar foils, specifically plurality-reversed forms of studied items, were used on a recognition test as opposed to the standard paradigm that uses unrelated foils in the recognition test. Furthermore, they repeated the presentations of the studied words and then, at test, asked for judgments of frequency (JOFs). The mean JOF for items judged to be old (that is, JOF > 0) increased with repetitions for both targets and similar foils for both low- and high-frequency words, continuing the reversal of the false-alarm pattern with word frequency such that there were more false alarms for low-frequency similar words.

Heekyeong Park, Lynne M. Reder, and Daniel Dickison, Department of Psychology, Carnegie Mellon University.

This work was supported by Grant 2-R01-MH52808 from the National Institute of Mental Health to Lynne M. Reder. We thank Norbou Buchler, Rachel Diana, and Elisabeth Ploran for commenting on a draft of this article.

Correspondence concerning this article should be addressed to Heekyeong Park, Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213. E-mail: hkpark@andrew.cmu.edu

Malmberg et al. showed that both the word frequency effect and the pattern of higher false-alarm rates for low-frequency similar foils could be explained within the retrieving effectively from memory (REM; Shiffrin & Steyvers, 1997) framework. For the explanation of the word frequency mirror effect, in the REM model, it is assumed that the memory representations for low-frequency words have features that are less familiar, and people also consider low-frequency words less likely to be encountered (Malmberg & Murnane, 2002; Malmberg et al., 2002). Thus the matching of the features associated with low-frequency words leads to more evidence for "old" judgment than does the matching of the features associated with high-frequency words, which explains the hit rate advantage for low-frequency words. However, when foils are very similar to studied items, it is expected that the typical false-alarm advantage for low-frequency words over high-frequency words will be reversed, because low-frequency similar foils have more diagnostic features for matching (Malmberg, Holden, & Shiffrin, 2004; Shiffrin & Steyvers, 1997). In the REM model, familiarity with different diagnostic values of different frequency words produces higher hit rates and lower false-alarm rates for low-frequency targets and dissimilar foils but higher false-alarm rates for low-frequency similar foils. Moreover, Malmberg, Holden, & Shiffrin (2004) challenged whether the SAC model could account for their finding of an interaction between word frequency and item similarity. They argued that the recollection process of the SAC model should facilitate the discrimination of targets from similar foils and should have counteracted any increased familiarity due to their similarity.

Figure 1 provides a schematic illustration of how SAC represents plurality information in a study event. In SAC, different concept nodes for different forms of a concept including different plurality forms are not assumed. Rather, it is assumed that an additional association binds to the episode node when participants are aware of the need to code whether the word is in the plural or singular form, much the same way as we have postulated that font is represented when participants bother to encode font information (Reder, Donavos, & Erickson, 2002).

There are two ways to recognize a word according to SAC: one based on recollection of the episode through sufficient activation at the episode node and the other based on a familiarity process that

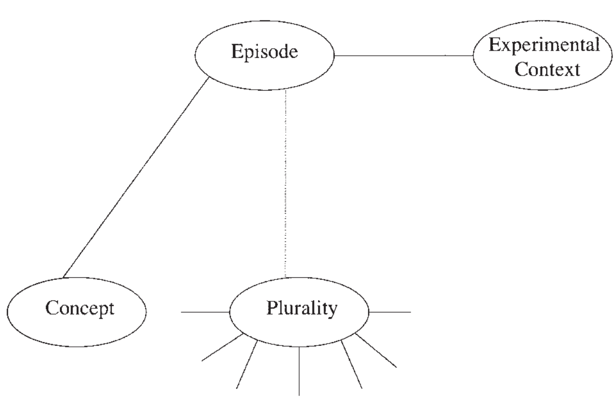

Figure 1. A schematic representation of words with plurality feature in the source of activation confusion model.

evaluates whether there is sufficient activation at the concept node to judge that the word had been studied in the experiment. According to SAC, most false alarms occur when "old" judgments are based on the error-prone familiarity process. However, false alarms can occur when the probe is similar to the studied word and when a critical feature forming the basis of the discrimination is not represented during the study phase. When participants do not appreciate the importance of coding plurality information during the study phase, they are much less likely to represent it as part of the episodic trace. At test, participants could try to retrieve the plurality information; however, if that information was not encoded during the study phase, these efforts are unsuccessful, resulting in recollections that may be inaccurate. Then, false alarms based on recollection of the similar word will occur more often for low-frequency words than high-frequency words, although false alarms based on familiarity will still be greater for high-frequency words.

In the situation in which participants must discriminate a plurality reversed foil from the studied form without knowing to code for that feature during study, low-frequency similar foils are more vulnerable to spurious recollection, for the reasons given above. In contrast, in the REM model, the higher false-alarm rates for low-frequency words are attributed to the notion that low-frequency words are more diagnostic for matching in recognition. Therefore, similar foils for low-frequency words still have more diagnostic matching features and are more likely to pass threshold for an "old" response than are high-frequency words with fewer diagnostic features. Although Malmberg, Holden, and Shiffrin (2004) introduced a "recall-like process" for the REM model, their explanation for higher false-alarm rates for low-frequency similar foils largely depends on the familiarity process. In contrast, in SAC, it is posited that the recollection process is responsible for the higher false-alarm rates for low-frequency similar foils.

In the present study, we tested a mechanistic account of what underlies the false-alarm patterns for different normative frequency words. We varied whether participants were warned prior to the study phase that the recognition test would require rejection of foils that differed only by a specified feature (font or plurality) from a studied word. In the other aspect of the study, we sometimes allowed participants to accept similar foils, but at the same time asked them to distinguish "old" judgments for which they had a recollection of a study event from those based on familiarity using the labels *remember* and *know* for those two classes, respectively (Tulving, 1985). Researchers have used this paradigm to investigate the contribution of different memory processes to recognition, although a few recent authors have questioned whether remember versus know judgments actually discriminate between single- and dual-process models of recognition memory (e.g., Dunn, 2004; Malmberg, Zeelenberg, & Shiffrin, 2004).[1]

[1] It is beyond the scope of this article to delve into the question of whether the remember–know paradigm is inconsistent with a signal detection theory interpretation (see Dobbins, Khoe, Yonelinas, & Kroll, 2000; Dunn, 2004; Yonelinas, 2002, for details). Dunn (2004) argued that the functionally dissociated remember–know data can also be handled with an equal variance model, but he did not address word frequency effects.

At issue is whether a dual-process mechanism can account for word frequency differentially affecting false alarms for similar foils. In particular, we tested whether the SAC mechanism that explains this reversal in word frequency false alarms arises from the recollection of an incomplete encoding trace. By manipulating the participants' knowledge of the importance of a specific feature, we expected to affect the content of the encoding trace. When no warning about the critical feature was given, we expected higher false-alarm rates for low-frequency similar foils. The typical word frequency mirror effect should still have occurred for targets and dissimilar new items. We expected more false alarms for similar foils of low-frequency words than for high-frequency words because retrieval of the encoding traces for low-frequency concepts is easier than for their high-frequency counterparts. In SAC, it is assumed that more activation arrives at episode nodes associated with low-frequency words. Low-frequency words have fewer prior contextual associations, and therefore there is less competition for activation from low-frequency concept nodes than there is from high-frequency concept nodes.

On one hand, the tendency to false alarm on the basis of erroneous recollections depends not only on the probability of sufficient activation arriving at the relevant episode node but also on the likelihood that the episode node has represented the critical feature that was varied at test. The likelihood of that feature being represented as part of the episode node is affected by whether the participant was warned during encoding. On the other hand, when participants are advised to attend to the critical feature of the study words, this greater tendency to false alarm to similar foils for low-frequency words should disappear. To the extent that participants recall the encoding event and have stored the critical feature as part of the memory trace, they should be more accurate at verifying whether the target is identical to the original studied item or a related foil. Retrieving the encoding trace should be easier for low-frequency words.

Therefore, when retrieval of the encoding trace includes the critical feature information, low-frequency words should outperform high-frequency words, whereas low-frequency words should be more vulnerable when that information was not encoded. As an additional test of this explanation, we employed remember–know judgments: if the reversed false-alarm pattern for low- and high-frequency words is based on the recollection process, then we should have observed these effects mainly in the remember responses. We tested these predictions using two types of critical features: plurality and font. This allowed us to manipulate within subject whether a critical feature was made salient prior to the study of the words on the list.

Method

Participants

Thirty-six undergraduate students at Carnegie Mellon University participated in return for course credit or for a payment of $10. All participants were native English speakers.

Task Overview

Table 1 lists the order of the experimental manipulations. We tested participants individually, and they completed four study–test cycles. The first two lists varied either the plurality of studied words or the font of these words. The last two lists varied the feature that was not varied for the first two lists. We fixed the order of study instructions (no-warning vs. warning about a critical feature that might be varied at test) and type of test in terms of discrimination required (discriminate vs. not to discriminate) to avoid carryover effects. To be specific, study instructions before the presentation of the first list did not mention anything about a critical feature presented during study or the importance of the critical feature for a later test. The first test also did not ask participants to discriminate between variants of studied words (plurality reversed or swapped font), so that participants would not be particularly motivated to encode those features for the second study list. Whenever participants were not asked to discriminate between variants of the studied word (Lists 1 and 4), they were instead asked to discriminate among their "old" judgments, specifically whether they remembered seeing the word or just felt that they "knew" that the word was old even though they had no specific recollection. No mention was made that they should give a remember response if the probe was an exact match and "know" otherwise. In this way, we could get an estimate of the extent to which accepting a similar word was based on recollection.

The second study-test list was similar to the typical plurality reversal experiment in that participants were not warned about a varied feature at study but were required to discriminate the studied form from similar foils. Knowledge that features could vary between study and test and that discriminating between the two versions mattered was established after List 2 and therefore, for the remaining two lists, participants were advised to attend to the critical feature that would vary. The feature that varied for the third and fourth lists was different from the feature used for the first two lists (plurality or font) in order to minimize practice at encoding specific features. Although the fourth list also warned participants to pay attention to the critical feature, at test, they were surprised in that they were told to ignore whether the feature varied from encoding and just respond on the basis of whether the concept had been studied during study.

Lists 1 and 2 differed from Lists 3 and 4 in that participants were not warned at study that the feature mattered for the first two lists. Lists 1 and 4 also differed from Lists 2 and 3 in that neither List 1 nor List 4 required a discrimination between the studied item and similar item. As mentioned above, both Lists 1 and 4 also asked participants to make remember–know–new judgments. For Lists 2 and 3, participants were asked to

Table 1
The Order of Experimental Manipulations

List	Study instructions	Test type	Critical feature
1	No warning	R–K–N (discrimination not required)	Font (or plurality)
2	No warning	O–N (discrimination required)	Font (or plurality)
3	Warning	O–N (discrimination required)	Plurality (or font)
4	Warning	R–K–N (discrimination not required)	Plurality (or font)

Note. R–K–N = remember–know–new recognition; O–N = old–new recognition.

discriminate similar foils from studied items and just made old–new recognition judgments. Finally, Lists 2 and 4 contained an element of surprise. Prior to the test for List 2, participants were unaware that they would be responsible for discriminating studied words from similar foils. Prior to the test for List 4, participants thought that they would be responsible for this type of discrimination but were then asked to ignore that feature at test.

Design, Materials, and Procedure

Whether the participant was warned prior to study about a critical feature of the stimulus (warning vs. no warning), word frequency (low vs. high), critical feature (plurality vs. font), type of test probe (target vs. similar vs. new), and recognition test (old–new-discriminate vs. remember–know–new-not to discriminate) was varied in a within-subject design with a fixed order of conditions; however, we manipulated whether a particular feature (plurality or font) was used for the first two tests or for the latter two between subjects.

Each test list consisted of studied words (target), similar foils that were plurality-reversed or swapped font of a studied word (similar), and unstudied words (new). Words assigned to target and similar item types were presented in a study list, but words assigned to the new item type were presented only in the test list. Four lists of words were constructed for each participant with the constraint that all words were between 4 and 10 letters in length. Half of the items in each list were high-frequency words, and the other half were low-frequency words, yielding a test list length of 96 items. Low-frequency items occurred fewer than 4 times per million words, and high-frequency items occurred more than 40 times per million (Kučera & Francis, 1967). Assignment of words to each condition was determined randomly for each participant.

For font manipulation, we used Comic Sans MS Bold and Georgia Bold Italic fonts. Figure 2 shows examples of words in the two fonts. When plurality was used as the critical feature, half of the words in a study list were presented in singular form, the other half were presented in plural form, and all words were presented in Times Roman font. When font was the critical feature, half of the words were presented in Comic Sans MS Bold font, the other half were presented in Georgia Bold Italic font, and all words were presented in singular form. When plurality was manipulated in the warning condition, font was manipulated in the no-warning condition and vice versa.

In the no-warning condition (Lists 1 and 2), participants were simply asked to encode the words as best they could. In the warning conditions (Lists 3 and 4), participants were told that they would see a series of words differing in a critical feature and that they would be responsible for distinguishing whether the test item matched the critical feature of the study item. For the plurality feature variations, participants were told that generating an image would help to discriminate whether the word was singular or plural. For the font feature variations, they were encouraged to judge how appropriate the font seemed for the meaning of the word. During study, words were presented one at a time for 1.5 s each. Participants made recognition judgments at their own pace. At the completion of the last test list, the participants were debriefed on the purpose of the experiment.

Results

The results were analyzed separately for the lists that asked for old–new judgments based on discriminating similar foils from old

ghoul ***flea***
room ***part***

Figure 2. An example of the fonts used at study.

items (Lists 2 and 3) and the lists that did not require this discrimination but did ask for remember–know–new judgments (Lists 1 and 4). Both font and plurality showed a Warning × Word Frequency interaction for similar items on old–new recognition such that, with no warning, there were more false alarms to low-frequency similar foils, font: low frequency (LF) = .63, high frequency (HF) = .40; plurality: LF = .50, HF = .44, but the difference disappeared with warning, font: LF = .33, HF = .35; plurality: LF = .10, HF = .16. This same pattern was found for the remember responses for both font and plurality, no-warning font: LF = .52, HF = .23; no-warning plurality: LF = .66, HF = .44; warning font: LF = .36, HF = .24; warning plurality: LF = .53, HF = .45. Because different participants got different features in different conditions, we collapsed over features in order to treat the experiment as a within-subject design. We used repeated measure analyses of variance with an alpha level of .05 for all statistical tests. Given that the Malmberg, Holden, and Shiffrin (2004) study involved old–new recognition judgments (that is, JOF >1), we discuss those results first.

Old–New Judgments

The mean hit rates for old items and the mean false-alarm rates for similar and new items in the no-warning (List 2) and warning (List 3) conditions are presented in Figure 3. Separate analyses for List 2 showed that a main effect of word frequency for similar foils was replicated as in Malmberg, Holden, and Shiffrin (2004) such that participants tended to make greater false alarms to low-frequency similar foils, $F(1, 35) = 14.45$, $MSE = .026$, $\eta^2 = .29$, with the word frequency mirror effect, $F(1, 35) = 10.42$, $MSE = .011$, $\eta^2 = .23$. For both Lists 2 and 3, we found the typical word frequency mirror effect, $F(1, 35) = 26.79$, $MSE = .011$, $\eta^2 = .43$, with more hits, $F(1, 35) = 9.22$, $MSE = .012$, $\eta^2 = .21$, but fewer false alarms, $F(1, 35) = 22.86$, $MSE = .008$, $\eta^2 = .39$, for low-frequency than for high-frequency words.

Overall, there was a main effect of warning such that participants made more "old" responses when they were not warned about importance of the critical feature, $F(1, 35) = 13.55$, $MSE = .024$, $\eta^2 = .28$, and this effect was primarily driven by the similar foils, $F(1, 35) = 39.25$, $MSE = .061$, $\eta^2 = .53$. There were significant Warning × Word Frequency, $F(1, 35) = 11.23$, $MSE = .016$, $\eta^2 = .24$, Warning × Type of Probe, $F(2, 70) = 29.32$, $MSE = .039$, $\eta^2 = .46$, Word Frequency × Type of Probe, $F(2, 70) = 13.64$, $MSE = .014$, $\eta^2 = .28$, and Warning × Word Frequency × Type of Probe interactions, $F(2, 70) = 5.75$, $MSE = .012$, $\eta^2 = .14$. It is the last three-way interaction that supports the contention that the greater false-alarm rate for low-frequency than high-frequency similar foils occurs only when participants are not instructed that they should attend to the feature that will subsequently be used as a basis for discriminating old from new test probes.

Separate analyses for false alarms to similar items showed an effect of frequency such that more "old" responses were made to low-frequency similar items, $F(1, 35) = 5.18$, $MSE = .02$, $\eta^2 = .13$; however, there was a still stronger Warning × Word Frequency interaction effect, $F(1, 35) = 16.44$, $MSE = .018$, $\eta^2 = .32$, such that greater false alarms for low-frequency similar items did not occur when participants were warned to encode the critical

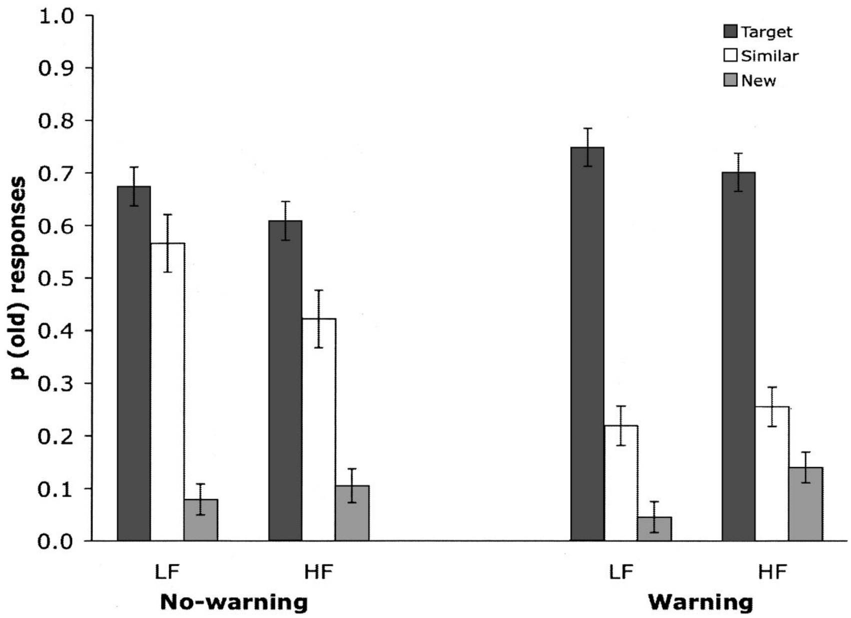

Figure 3. Hits to target items and false alarms to similar and new items on old–new recognition in the no-warning and warning conditions as a function of word frequency. LF = low frequency; HF = high frequency. Error bars depict 95% confidence intervals.

feature during study. This Warning × Word Frequency interaction was found only for similar foils and not for targets or new items (both *p*s $> .1$). For target items, more hits were found in the warning condition, $F(1, 35) = 7.31$, $MSE = .033$, $\eta^2 = .17$, but there was no warning effect to new items ($F < 1$).

Discrimination of targets from similar items (d') showed a main effect of warning, $F(1, 35) = 28.40$, $MSE = 2.19$, $\eta^2 = .45$, such that participants were better at discriminating targets from similar foils when given a warning. This effect also interacted with frequency, $F(1, 35) = 8.72$, $MSE = .289$, $\eta^2 = .20$, such that d' scores were lower for low-frequency words than for high-frequency words in the no-warning condition (.34 vs. .54), but that pattern was reversed in the warning condition (1.92 vs. 1.60).

Remember–Know–New Judgments

The proportions of remember versus know versus new judgments in the no-warning and warning conditions for Lists 1 and 4 are presented in Figure 4. Note that for these lists, participants were not required to judge similar items as new. For List 1 (no-warning), they were not made aware of the variability in features; however, for List 4 (warning), they expected to be tested on the distinction but told at test to accept similar items as if they had been studied.

Ignoring the similar items, and by collapsing remember and know responses into "old" responses, we replicated the standard word frequency mirror effect, $F(1, 35) = 82.83$, $MSE = .015$, $\eta^2 = .70$, such that there were more hits to low-frequency than there were to high-frequency targets, $F(1, 35) = 31.18$, $MSE = .014$, $\eta^2 = .47$, but fewer false alarms to low-frequency than to high-frequency foils, $F(1, 35) = 49.84$, $MSE = .018$, $\eta^2 = .59$. We found neither a warning main effect nor an interaction effect between warning and frequency for studied and new items.

Participants gave more remember responses for low-frequency targets, $F(1, 35) = 48.95$, $MSE = .024$, $\eta^2 = .58$, and more know judgments for high-frequency targets, $F(1, 35) = 13.35$, $MSE = .014$, $\eta^2 = .28$, consistent with the pattern found in Reder et al. (2000). Significantly more of the false alarms given as know responses were for high-frequency new words than for low-frequency words, $F(1, 35) = 54.62$, $MSE = .009$, $\eta^2 = .61$; however, remember responses for new items did not reliably differ as a function of word frequency or warning.

On the other hand, the analysis for similar items showed a main effect of word frequency for remember responses such that participants made more remember responses for low-frequency similar items, $F(1, 35) = 60.61$, $MSE = .018$, $\eta^2 = .63$. This effect was modulated by an interaction effect with warning, $F(1, 35) = 8.87$, $MSE = .024$, $\eta^2 = .20$, such that remember responses for low-frequency similar items decreased with warning, $F(1, 35) = 4.56$, $MSE = .08$, $\eta^2 = .12$, but remember responses for high-frequency similar items did not change with warning ($F < 1$). We found no main effect due to warning.

Know responses to similar items yielded a main effect of word frequency, $F(1, 35) = 4.92$, $MSE = .013$, $\eta^2 = .12$, such that high-frequency items gave more know responses. There was also a Warning × Word Frequency interaction, $F(1, 35) = 5.27$, $MSE = .011$, $\eta^2 = .13$, such that participants made more know responses to high-frequency similar items in the no-warning con-

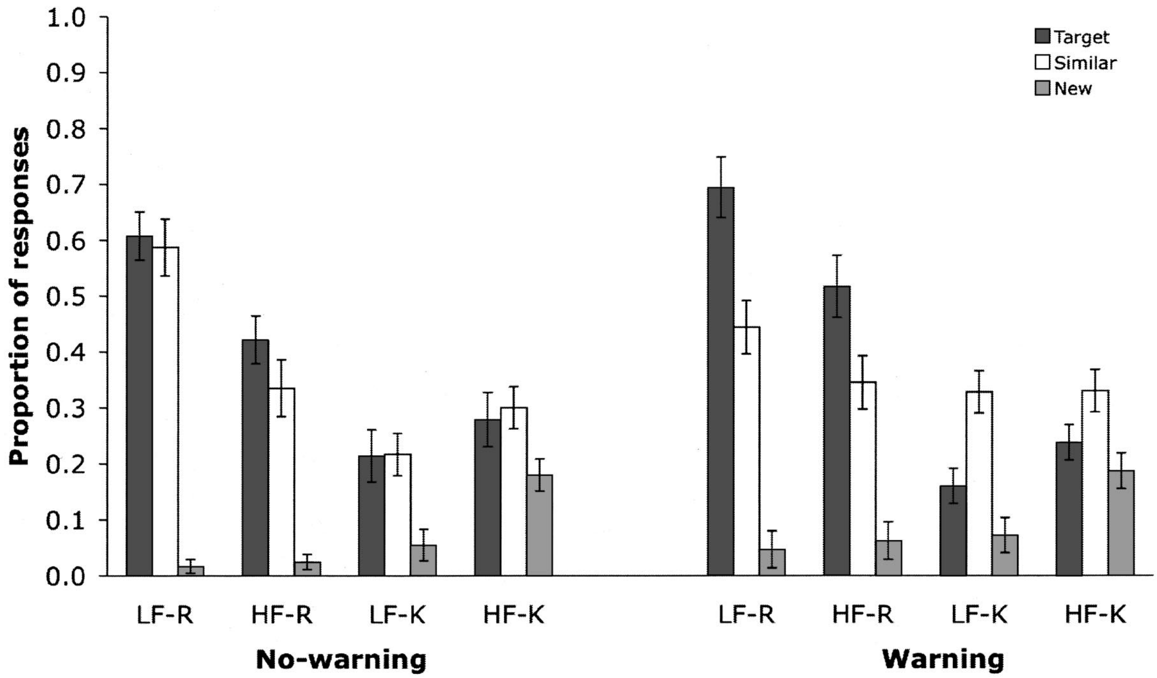

Figure 4. The proportions of remember, know and new judgments as a function of warning, word frequency and item type. LF-R = remember response to low-frequency item; LF-K = know response to low-frequency item; HF-R = remember response to high-frequency item; HF-K = know response to high-frequency item. Error bars depict 95% confidence intervals.

dition, $F(1, 35) = 10.18$, $MSE = .012$, $\eta^2 = .23$, but not in the warning condition ($F < 1$). A familiarity measure [K/(1−R)] suggested by Yonelinas and colleagues (e.g., Yonelinas & Jacoby, 1996; Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1998) did not show a reliable difference in familiarity responding for low- and high-frequency words.

In the no-warning condition, participants were not asked to discriminate between targets and similar items. Nonetheless, one can calculate the sensitivity to the difference by treating original items as hits and similar items as foils. Discriminability indexed by d' for remember responses was .027 for low-frequency words, compared with .243 for high-frequency items. In contrast, when warned prior to encoding that the feature mattered for recognition even though participants were explicitly told not to distinguish between similar forms at test, discriminability nonetheless improved: d' scores for remember responses increased to .932 for low-frequency items and .734 for high-frequency items. This pattern produced a main effect of warning, $F(1, 35) = 13.34$, $MSE = 1.316$, $\eta^2 = .28$, but the Warning × Word Frequency interaction was not significant ($F < 4$).

Discussion

This study replicated previous results in terms of reproducing the word frequency mirror effect and also produced more false alarms to low-frequency similar items than to high-frequency similar items, as found by Malmberg, Holden, and Shiffrin (2004). In both our account and in the Malmberg et al. account, it was assumed that potentially mismatching features would be encoded as part of the representation. What makes our model different from theirs is that we assumed that the likelihood of storing and thus retrieving these potentially mismatching features would vary with encoding instructions. It is important to note that we demonstrated that encoding instructions could diminish the word frequency mirror reversal by reducing the number of false alarms to similar foils, in particular, the relative vulnerability of low-frequency similar foils. The instructional effect on recognition observed in this study is not always obtained and appears to be affected by the nature of the encoding instructions, not the warning per se.[2] This study not only informed participants to encode the critical feature for a later test but also provided specific examples designed to help encode the critical feature.

Another difference between our account and that of Malmberg, Holden, and Shiffrin (2004) is that we assumed that the reversed

[2] The effect observed here but not always obtained may also have been due to our using an old–new recognition task instead of a JOF task and because there was no repetition of items (e.g., Cleary, Curran, & Greene, 2001; Hintzman & Curran, 1995; Sheffert & Shiffrin, 2003). That meant that the list length was relatively short and that each item in our study was less disadvantaged than once-presented items that are competing with items repeated multiple times in a list (e.g., Diana & Reder, in press; Norman, 2002).

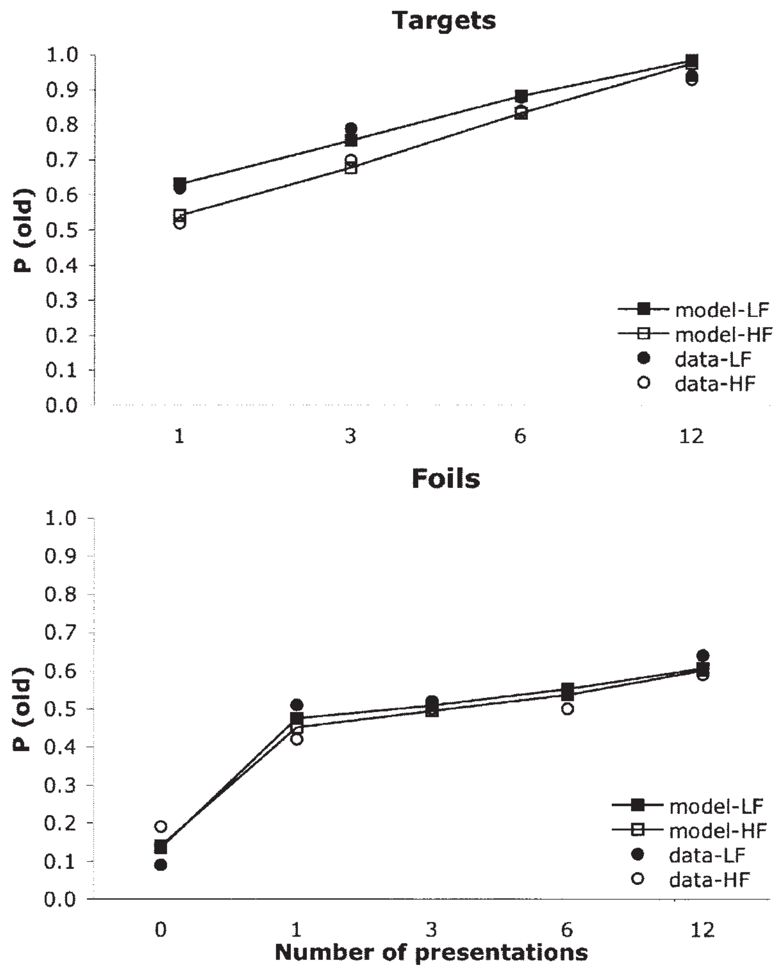

Figure 5. Source of activation confusion model predictions for probability of responding "old" as a function of the number of presentations (data from Malmberg, Holden, & Shiffrin, 2004). LF = low frequency; HF = high frequency.

mirror effect derived from the fact that when these mismatching features are not retrieved at test, the spurious "old" responses are likely to be based on a faulty recollection rather than a familiarity response. The test lists that required participants to give remember–know responses provide evidence to support the view that the greater acceptance of similar items for low-frequency words was mainly based on the greater ease of recollecting low-frequency words. In the no-warning condition, there were more remember judgments for low-frequency similar items than there were for high-frequency similar items. This pattern is consistent with the SAC account that low-frequency words can more easily access an episodic trace that affords recollection. The SAC explanation is based on the assumption that there are more prior contextual associations to high-frequency words, and that more contextual associations translates into greater interference in retrieving any specific context such as the relevant one during the experiment (see Reder et al., 2000, for more details). This assumption is formalized as less activation spreading from the concept to the relevant episode node, because this activation is spread across more associations for a high-frequency word.

The effect of recollection in recognition of similar foils has been supported by other studies. Rotello and Heit (2000) found evidence for recall-to-reject processing in associative recognition. Rotello, Macmillan, and Van Tassel (2000) also demonstrated that recall-to-reject processing was involved in plurality recognition when participants were explicitly instructed that plurality-reversed foils would be present. However, the use of the recall-to-reject process declined when instructions were more vague. These data seem consistent with our perspective.

Another assumption adopted by the SAC model of recognition is that familiarity as well as recollection can be used to make an "old" response. We assume that familiarity-based judgments come from the activation level of the concept node rather than the episode node. High-frequency words not only have more preexperimental contextual fans but also have a higher familiarity value, and both are due to a greater prior history of exposure. This implies

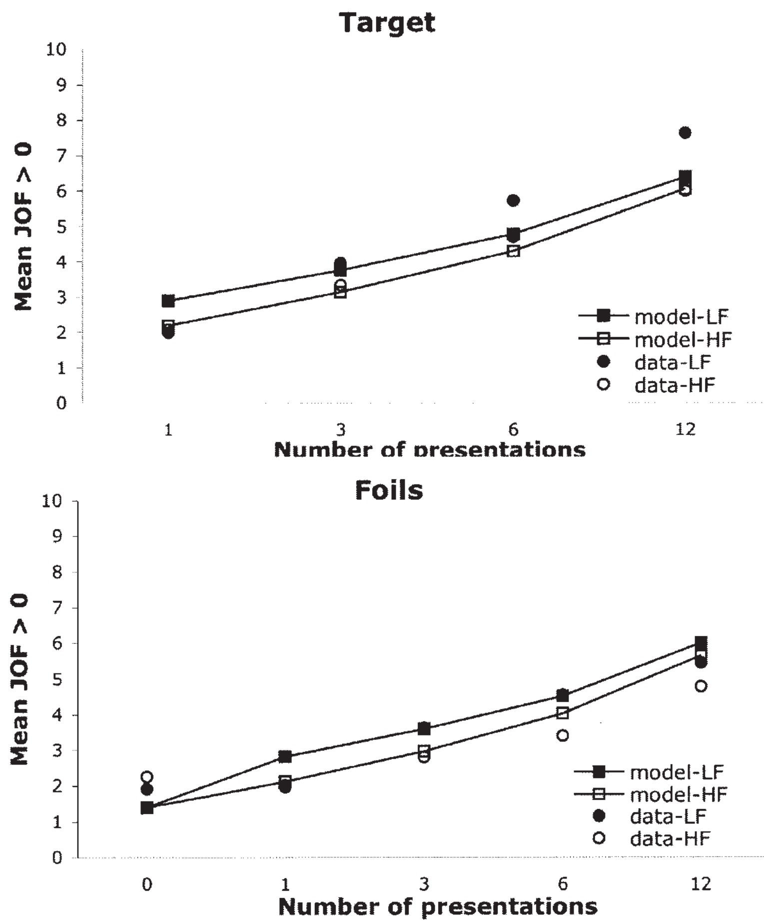

Figure 6. Source of activation confusion model predictions for judgments of frequency as a function of the number of presentations (data from Malmberg, Holden, & Shiffrin, 2004). LF = low frequency; HF = high frequency.

that there should be more familiarity-based false alarms for high-frequency new items because they have a higher resting level of activation, making them more likely to exceed a familiarity threshold.

When participants were not aware that the plurality or font of the word was a critical feature of the study item, this critical feature might not have been especially well bound to the encoding trace. With warning, participants became more accurate, presumably because they attended to the relevant feature and associated it with the episode node. According to SAC, the remember responses should have increased slightly for the exact matches in the warned condition. With an exact match, there is an additional context source from which activation can spread to the episode node. Because this type of context has such high fan, the increase in activation is expected to be small. It has been demonstrated that the value added by matching the additional contextual source is modulated by the number of words sharing that context (Cary & Reder, 2003; Reder et al., 2002). When activation exceeds threshold on the episode node, participants can attempt to retrieve the critical feature that may have been bound to the episode node. However, features that are not salient and were not made salient from instructions are often not encoded. Furthermore, the number of exposures to the stimulus may not dramatically change the probability that the critical feature is encoded if that feature is not treated as important.

The notion of incomplete encoding processes is hardly a new idea. Nickerson and Adams (1979)'s demonstration that Americans are poor at discriminating an accurate drawing of a penny from similar foils illustrates that people can experience something as common as a penny thousands of times without encoding all of the features that enable discrimination of the target from similar distractors. On the other hand, when motivated to learn a discrimination such as in the warning condition of this study, participants are more likely to have encoded the critical feature, making recognition more veridical.

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

Figure 5 presents the fit of a SAC model to the empirical results of Malmberg, Holden, and Shiffrin (2004) for "old" responses. The top panel displays the fit for targets as a function of normative word frequency and the number of experimental presentations. The bottom panel displays the fit for false alarms. Figure 6 shows the model's fit for JOFs that are greater than zero in an analogous fashion to Figure 5. The theoretical account of the word frequency mirror effect using the SAC framework and the model simulation with remember–know responses are found in Reder et al. (2000), along with the parameters and equations for the model fit of the word frequency effect. In addition to using previously estimated parameters and equations from previous model fits (see Cary & Reder, 2003; Reder et al., 2000; Schunn, Reder, Nhouyvanisvong, Richards, & Stroffolino, 1997, for more details of the SAC model), we used new equations used to fit the Malmberg et al. data, and these are presented in the Appendix along with the parameter values that were estimated to optimize the fit.

It is assumed that the probability of encoding the critical feature is low when participants are not warned that this information is important. The probability of an "old" response depends on the probability of a recollection plus the probability of a know response when recollection fails. The probability of a recollection is determined by how the activation value that accrues at the episode node compares to the retrieval threshold, assuming a normal distribution of activation values and the standard deviation of the episode node also estimated. The amount of activation at an episode node depends on its previous level of activation (from number of times the event was experienced and how long ago it had been experienced) and the amount that spreads to it.

We adjusted the probability estimate of a recollection to exclude those cases in which the episode node was retrieved but the feature information was also retrieved and mismatched the probe. The probability of finding such a mismatch is relatively small, because it depends on the probability of actually having formed that link and of enough activation having accrued to get the node over its threshold. The mean JOFs > 0 were estimated from the activation values accrued at the episode node. The model fits showed that SAC is capable of fitting the probability of responding "old" for an item and mean JOF > 0 reported in the Malmberg, Holden, and Shiffrin (2004) study, recognition–target: $R^2 = .99$, root-mean-square deviation (RMSD) = .03; recognition–foil: $R^2 = .98$, RMSD = .03; JOF–target: $R^2 = .97$, RMSD = .66; JOF–foil: $R^2 = .95$, RMSD = .56.

In summary, we showed how a dual-process model such as SAC can explain the reversed word frequency mirror effect for similar foils. Specifically, we demonstrated that the higher false alarms for low-frequency similar items come from the recollection process. In addition, we showed that this higher false-alarm rate for low-frequency words disappears when participants are aware that the feature is important during study. In addition to providing an account and evidence for the reversed word frequency mirror effect for similar items within the SAC framework, we also demonstrated a good quantitative fit to the Malmberg, Holden, and Shiffrin (2004) data. The fits required few additional assumptions to the original SAC model of the word frequency effect.

References

Arndt, J., & Reder, L. M. (2002). Word frequency and receiver operating characteristic curves in recognition memory. Evidence for a dual-process interpretation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 830–842.

Balota, D. A., Burgess, G., C., Cortese, M. J., & Adams, D. R. (2002). Memory for the infrequency 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.

Cary, M., & Reder, L. M. (2003). A dual-process account of the list-length and strength-based mirror effects in recognition. *Journal of Memory and Language, 49,* 231–248.

Cleary, A., Curran, T., & Greene, R. L. (2001). Memory for detail in item versus associative recognition. *Memory & Cognition, 29,* 413–423.

Diana, R. D., & Reder, L. M. (in press). The list strength effect: A contextual competition account. *Memory & Cognition.*

Dobbins, I. G., Khoe, W., Yonelinas, A. P., & Kroll, N. E. (2000). Predicting individual false alarm rates and signal detection theory: A role for remembering. *Memory & Cognition, 28,* 1347–1356.

Dunn, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review, 111,* 524–542.

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

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 5–6.

Guttentag, R. E., & Carroll, D. (1997). Recollection-based recognition: Word frequency effects. *Journal of Memory and Language, 37,* 502–516.

Hintzman, D. L. (1994). On explaining the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 201–205.

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

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

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.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present day American English.* Providence, RI: Brown University Press.

Malmberg, K. J., Holden, J., & Shiffrin, R. M. (2004). Modeling the effects of repetitions, similarity and normative word frequency on old-new recognition and judgments of frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 319–331.

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., 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,* 540–549.

Murdock, B. B. (1998). The mirror effect and attention-likelihood theory: A reflective analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 524–534.

Murdock, B. B. (2003). The mirror effect and the spacing effect. *Psychonomic Bulletin & Review, 10,* 570–588.

Nickerson, R. S., & Adams, M. J. (1979). Long-term memory for a common object. *Cognitive Psychology, 11,* 287–307.

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

Reder, L. M., Angstadt, P., Cary, M., Erickson, M. A., & Ayers, M. S. (2002). A reexamination of stimulus-frequency effects in recognition: Two mirrors for low- and high-frequency pseudowords. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 138–152.

Reder, L. M., Donavos, D. K., & Erickson, M. A. (2002). Perceptual match effects in direct tests of memory: The role of contextual fan. *Memory & Cognition, 30,* 312–323.

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., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition, 28,* 907–922.

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

Schunn, C. D., Reder, L. M., Nhouyvanisvong, A., Richards, D. R., & Stroffolino, P. (1997). To calculate or not to calculate: A source activation confusion model of problem familiarity's role in strategy selection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 3–29.

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

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

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26,* 1–12.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46,* 441–517.

Yonelinas, A. P., & Jacoby, L. L. (1996). Noncritical recollection: Familiarity as automatic, irrelevant recollection. *Consciousness and Cognition, 5,* 131–141.

Yonelinas, A. P., Kroll, N. E., Dobbins, I., Lazzara, M., & Knight, R. (1998). Recollection and familiarity deficits in amnesia convergence of remember-know, process dissociation, and receiver operating characteristic data. *Neuropsychology, 12,* 323–339.

(*Appendix follows*)

Appendix

SAC Model Parameters

Parameter	Function	Value
T_c	Concept node decision threshold	300.00
T_e	Episode node decision threshold	33.72
T_f	Feature node decision threshold	3495.20
σ_c	Concept node standard deviation	233.83
σ_e	Episode node standard deviation	87.66
σ_f	Feature node standard deviation	20.13
σ_{JOF}	JOF standard deviation	1.37
C_{JOF}	Constant for JOF	0.44
P_f	Probability that a link exists between episode node and feature node	0.38

SAC Model Equations

New equations	Description
$A_f = B_f + \left(\frac{A_e S_{e,f}}{\Sigma S_e}\right)$	Activation of feature node for recollection decision where B_f is the base level activation of feature node, A_c is activation of episode node, $S_{e,f}$ is link strength between episode node and feature node, and ΣS_e is the sum of the strengths of all links from episode node
$P(f) = N[(A_f - T_f)/\sigma_f]$	Probability of feature recollection
$P(R_{FA}) = P(e)[1 - P_f P(f)]$	Probability of remember response for false alarms, where $P(e)$ is the probability that the episode node activation is above threshold
$P(j = 0) = 1 - [P(R) + P(K)]$	Probability of responding JOF = 0 where $P(K)$ is the probability of know response
$P(j > 0) = s \times \{N[((j + 1) - M_{JOF})/\sigma_{JOF}] - N[(j - M_{JOF})/\sigma_{JOF}]\}$ $s = \left(\frac{1 - P(j = 0)}{1 - N[(1 - M_{JOF})/\sigma_{JOF}]}\right)$	Probability of responding JOF > 0 where mean of JOF $N[x]$ is the area under the distribution curve to the left of x with mean of JOF ($M_{JOF} = C_{JOF}\sqrt{A_e}$) and standard deviation (σ_{JOF})

SAC model equations for the word frequency effect (Reder et al., 2000)	Description
$B = B_w + c_N \Sigma t_i^{-dN}$	Base-level activation as a function of delay and repetitions
$\Delta A = -\rho(A - B)$	Change in current activation from one trial to the next
$\Delta A_r = \Sigma(A_s S_{s,r}/\Sigma S_{s,i})$	Change in current strength because of activation spread
$S_{s,r} = c_t \Sigma t_i^{-dL}$	Link strength as a function of delay and repetitions
$P(R) = P(e)$	Probability of responding R as a function of current activation, in which $P(e) = N[(A_e - T_e)/\sigma_e]$
$P(K) = [1 - P(e)]P(c)$	Probability of responding K as a function of current activation, in which $P(c) = N[(A_c - T_c)/\sigma_c]$

Note. SAC = source of activation confusion; JOF = judgment of frequency.

Received July 26, 2004
Revision received November 16, 2004
Accepted November 24, 2004 ■