

# Not all sources of familiarity are created equal: the case of word frequency and repetition in episodic recognition

Jennifer H. Coane · David A. Balota ·  
Patrick O. Dolan · Larry L. Jacoby

© Psychonomic Society, Inc. 2011

**Abstract** Low-frequency (LF) words produce higher hit rates and lower false alarm rates than high-frequency (HF) words. This word frequency mirror pattern has been interpreted within dual-process models of recognition, which assume the contributions of a slower recollective process and a relatively fast-acting familiarity process. In the present experiments, recollection and familiarity were placed in opposition using Jacoby, L. L., *Journal of Memory and Language*, 30, 513–541 (1991), two-list exclusion paradigm with HF and LF words. Exclusion errors to LF words exceeded those to HF words at fast deadlines, whereas the reverse occurred at slow deadlines. In [Experiments 2](#) and [3](#), false alarms to HF nonpresented lures were higher than to LF nonpresented lures, indicating the use of baseline familiarity for totally new items. Combined, these results indicate that in addition to baseline familiarity and recollection, a third process involving the assessment of a relative change in familiarity is involved in recognition performance. Both relative changes in familiarity and recollection processes have distinct time courses and are engaged when there is diagnostic list information available, whereas baseline

familiarity is used when there is no diagnostic information available (e.g., for nonpresented lure items).

**Keywords** Word frequency · Recognition · Recollection · Relative and absolute familiarity · Mirror effect

The concept of familiarity is intuitively straightforward: To be familiar with something involves knowing that one has encountered it, even when one is unable to specify when or where the encounter took place. For example, at a professional conference, one might see a familiar face without being able to retrieve the name or even the context from which one knows the individual. The notion of familiarity plays a central role in dual-process models of recognition performance (e.g., Jacoby & Dallas, 1981; Yonelinas, 2002). In these models, familiarity is contrasted with recollection. *Recollection* refers to retrieving specific details about the item's prior occurrence. Source (e.g., was it visually or auditorily presented?) or temporal information (e.g., was it in the first or second list?) is characteristic of recollection, whereas familiarity assessments do not involve such information. When an item is presented in a recognition test, individuals can respond on the basis of one or both processes. For example, if specific details cannot be accessed at the time of retrieval (i.e., when recollection fails), one can use familiarity to correctly identify an item.

Dissociating the contribution of each process in the standard recognition test paradigm is challenging, because many manipulations can increase both recollection and familiarity, although the two processes may be differentially sensitive to such manipulations. For example, deeper levels of processing and spacing of presentations result in larger increases in recollection than in familiarity (e.g., Benjamin & Craik, 2001; Yonelinas, 2001), whereas changes in modality between study and test tend to influence familiarity more than recollection. Increasing study duration and

---

J. H. Coane (✉)  
Department of Psychology, Colby College,  
5550 Mayflower Hill,  
Waterville, ME 04901, USA  
e-mail: jhcoane@colby.edu

D. A. Balota · L. L. Jacoby  
Washington University in Saint Louis,  
St. Louis, MO, USA

P. O. Dolan  
Drew University,  
Madison, NJ, USA

item repetition increase both familiarity and recollection (e.g., Benjamin & Craik, 2001).

Although a full review of dual-process models is beyond the scope of this article (see Diana, Reder, Arndt, & Park, 2006; Yonelinas, 2002, for reviews), for the present purposes, it is important to note that the two processes appear to have different time courses. In particular, in untimed retrieval tasks, responses based on familiarity tend to be faster overall than responses based on recollection (Atkinson & Juola, 1973, 1974; Boldini, Russo, & Avons, 2004; Hintzman & Caulton, 1997; Mandler, 1980; but see Dewhurst, Holmes, Brandt, & Dean, 2006). Familiarity is assumed to be a relatively automatic, fast-acting process that does not require extensive engagement of attentional control. In contrast, recollection takes more time and involves more controlled processes (Atkinson & Juola, 1974; Boldini et al., 2004; Jacoby, 1991; Yonelinas & Jacoby, 1994, 1996). When response deadlines are imposed, deadlines of 1,000 ms or less tend to increase reliance on familiarity, whereas recollection appears to come online later, producing biphasic distributions of response latencies (e.g., McElree, Dolan, & Jacoby, 1999). Furthermore, different tasks rely to a greater or lesser extent on these two processes, with tasks requiring decisions based on specific event information (e.g., source or context) reflecting the influence of recollection more than do tasks that require a discrimination at a more general level (e.g., old/new decisions).

Although there is a substantial body of work examining how recollection and familiarity differ, relatively less attention has been devoted to characterizing the nature of familiarity. This is the question addressed in the present study. Specifically, one question is whether the fast-acting familiarity process is a unitary construct or whether different sources of familiarity have different influences on a global assessment of familiarity as it is generally defined. A goal of the present research is to systematically address how changes in familiarity due to two different factors influence performance on a recognition test and whether the observed effects require a more fine-grained analysis of familiarity. The distinction we will be making throughout this article is between preexperimental familiarity, operationalized as normative word frequency, and experimental or context-dependent familiarity, operationalized as the number of times items were presented during an encoding context.

### The word frequency effect in recognition

Low-frequency (LF) words produce higher hit rates and lower false alarm rates than high-frequency (HF) words (e.g., Balota & Neely, 1980; Glanzer & Adams, 1985; Joordens & Hockley, 2000; Reder et al., 2000). This pattern is referred to as the *word frequency mirror pattern*. In the

past decade, two dual-process accounts of the word frequency mirror effect have been proposed (Joordens & Hockley, 2000; Reder et al., 2000). Although these models differ in some important ways, both assume that a familiarity-based process and an item-specific recollective process contribute jointly to produce hit rates, because both greater familiarity and recollection will increase the probability of a correct “old” response in standard recognition tasks. According to Joordens and Hockley, recollection processing depends on distinctiveness, which in turn is a function of the number of preexperimental contexts in which an item has been experienced. LF items, having occurred in fewer contexts, are more distinctive and thus incur a benefit compared to HF items (see Adelman, Brown, & Quesada, 2006). According to Reder and colleagues (Reder et al., 2000; Reder, Paynter, Diana, Ngiam, & Dickison, 2007), recollection depends on the activation level of “episode nodes.” The episode node’s activation is a function of the activation of the word itself (the “word node”) and the context in which it was studied. Thus, recollection depends on the retrieval of specific information that contextualizes the source of an item’s activation level. LF items have fewer preexperimental connections emanating from the word node, and hence the episodic connection is relatively stronger, thereby producing the recollection advantage compared to HF words. In addition to a recollection process that benefits LF words over HF words, both models assume the use of a baseline familiarity process that results in more incorrect “old” responses to HF foils than to LF foils, because HF words have higher baseline familiarity. Thus, these models predict that, in most cases, false alarms to HF foils should be higher than false alarms to LF foils, because only baseline familiarity can drive responses to new items.

If the above dual-process accounts are correct, the LF advantage in correct rejections is due to lower baseline familiarity, and the LF hit rate advantage is due to better recollection. Consequently, it should be possible to alter the extent to which participants use baseline familiarity or recollection information. Balota, Burgess, Cortese, and Adams (2002) and Joordens and Hockley (2000) showed that one can eliminate and even partially reverse the LF advantage in recognition by increasing participants’ reliance on baseline familiarity. In both studies, imposing a response deadline decreased the hit rate advantage for LF words, while leaving the word frequency effect for lures relatively unaffected. These results are quite consistent with dual-process accounts of the word frequency effect, in which a fast and relatively automatic familiarity process and a slower, more resource-demanding recollective process act in concert to drive recognition decisions.

## Absolute and relative familiarity

The more frequently an individual processes an item, the more familiar that item becomes, and, as expected, there is a large correlation between subjective familiarity and frequency of occurrence in the language (see, e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004). Within the context of a study list, an item that is presented three times will have a greater level of absolute familiarity (and recollection) than an item presented once. Here, we are defining absolute familiarity as the given familiarity signal of an item at a specific point in time, with the assumption that a study event will increase an item's absolute familiarity. For foils in a recognition test, baseline familiarity and absolute familiarity are identical. However, some changes in familiarity are better described as *relative*, in the sense that the assessment of familiarity depends on the *change* in familiarity from baseline as a result of a study event. Thus, although both absolute familiarity and relative familiarity change as a function of a study event, the former is a simple reflection of the item's current familiarity signal, but the latter includes information about the amount of change. Because LF words have a lower baseline familiarity level than HF words, LF words receive a relatively greater boost in familiarity due to recent presentation than do HF words, which are fairly insensitive to a single additional processing episode. Indeed, there is clear evidence of such a frequency  $\times$  repetition interaction in the familiarity-based lexical decision task (e.g., Balota & Spieler, 1999). We assume that HF words will have a higher absolute familiarity signal than LF words even after study, as it is unlikely that a single study event could cause an LF word's overall familiarity to exceed the familiarity of an HF word, which has benefited from a lifetime of repeated exposure. Thus, absolute/baseline familiarity should be greater for HF words, but relative familiarity should be greater for LF words. Mandler (1980) suggested that the recognition advantage of LF words, compared to HF words, is due to the relatively greater boost in familiarity due to a study event. Note that in this case, the recognition advantage for LF words is due to a relative familiarity boost, not to recollection, as proposed by dual-process models (Jacoby & Dallas, 1981).

Although there is some appeal to the relative-familiarity account as a mechanism underlying the word frequency effect in recognition, this perspective clearly has some limitations. Specifically, this account has difficulty accommodating the advantage of LF foils over HF foils in correct rejections. Because there is no relative change in familiarity for items that were not presented, participants must be using some other source of information to reject distractors; that is, LF words have lower baseline familiarity than HF words and are further from a

response threshold. Thus, to incorporate the concept of a *relative* change in familiarity to account for the word frequency mirror effect, one must concede that there are potentially two distinct types or sources of familiarity (i.e., an absolute/baseline level and a relative change) that contribute to recognition decisions.

It is important to note that our conceptualization of relative familiarity is quite distinct from the dual-process concept of recollection, in that the computation of relative changes in familiarity does not require controlled processes, as recollection does. That is, according to the dual-process models reviewed above, the recollection benefit for LF words is due to greater source-specific information for LF than for HF words; however, the relative-familiarity mechanism makes no assertions regarding source-specific information. More importantly, the relative-familiarity account makes a distinct prediction regarding the time course of the processes involved in recognition. Specifically, both familiarity-based processes should be fast-acting, whereas the source-specific recollective process should be slower to develop. Therefore, at fast response deadlines, frequency-modulated absolute/baseline and relative familiarity are expected to drive responses, with the former contributing to performance on both old and new items, and the latter contributing only to performance on old items. A slower-acting recollection process, which is also frequency modulated, should come online later in processing (see Joordens & Hockley, 2000; Reder et al., 2000). Both the fast-acting relative familiarity and the slower-acting recollection component are likely to be stronger for LF than for HF items, whereas the baseline familiarity component should be stronger for HF words. Thus, the primary difference between the dual-process models proposed by Joordens and Hockley and by Reder et al. and the tri-process account proposed here is the inclusion of a fast-acting assessment of relative familiarity for old items in addition to the more traditional conceptualizations of (baseline) familiarity for new items and recollection for old items. Critically, because relative and baseline familiarity are expected to act in opposing directions as a function of word frequency (the former will be stronger for LF words and the latter for HF words), it should be possible to dissociate these two sources of familiarity empirically.

## The present study

To test the influence of preexperimental (absolute/baseline) and experimental (relative) familiarity on recognition performance, we used the Jacoby (1999) two-list exclusion paradigm, which allows one to examine familiarity when it is placed in opposition to recollection (see Table 1 for a schematic of the events in the present study). In this version

**Table 1** Sample study and test items and response classification scheme

Study Phase		Test Phase		Test Outcome	
List 1 Visual	List 2 Auditory	Visual	Correct Response	Participant Response	Response Type
animal	skillet	animal	new	old	Exclusion error
arbiter	mastery	love	old	old	Hit
home	present	party	new	old	False alarm
irony	love	shriek	new	new	Correct rejection
.	gadfly	gadfly	old	new	Miss
.	money	truth	new	new	Correct rejection
.	.	.	.	.	.
panorama	.	.	.	.	.
animal	.	.	.	.	.
water	.	.	.	.	.
irony	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
truth	.	.	.	.	.
irony	.	.	.	.	.
shriek	.	.	.	.	.
animal	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

of the two-list exclusion paradigm, participants study two lists, one presented visually and one presented auditorily. The visual items are presented once or three times during study. At test, participants are instructed to endorse the auditory items, while rejecting items from the visual list as well as totally new items. Thus, participants must be able to recollect the specific source (auditory or visual presentation). For the to-be-excluded items, the increased familiarity due to an earlier visual study event will be in conflict with the recollection-based exclusion instructions, so false alarms to these items (hereafter *exclusion errors*) provide a measure of familiarity-based responding when recollection fails. Furthermore, thrice-presented items should have a higher level of familiarity than once-presented items (see Jacoby, 1999).

Because prior work using the two-list exclusion procedure has demonstrated that familiarity-based responding increases when a speeded response deadline is imposed (i.e., more errors occur for thrice-presented items), whereas recollection-based responding increases when sufficient time is allowed for processing (i.e., thrice-presented items result in fewer errors; see Jacoby, Jones, & Dolan, 1998; McElree et al., 1999), we used multiple response deadlines to examine the relative contributions of these processes. An early, more automatic evaluation of relative familiarity should be active

during the earliest portion of the response (i.e., faster than 1,000 ms). It should also yield a pattern different from the attention-demanding recollection process, which should be more engaged at later points in the response window (i.e., after 1,000 ms). To narrowly titrate response latencies in the first experiment, we employed the rhythmic-cuing paradigm used by Balota et al. (2002). In this paradigm, described below, participants are given multiple temporal cues for when to respond, thus minimizing deviations from the targeted response times (RTs).

The question addressed here is whether word frequency (HF vs. LF) and repetition (one vs. three) can dissociate the two sources of familiarity in the Jacoby (1999) two-list exclusion paradigm. In Jacoby (1999), under response pressure, repetition increased errors, consistent with a fast-acting familiarity-based process. However, when sufficient time was allowed for a response, repetition decreased errors, consistent with a slower recollection-based process. According to dual-process accounts of the mirror effect, one might expect similar effects of frequency and repetition in exclusion errors. Specifically, HF and thrice-presented words should be more familiar than LF and once-presented words, and thus should result in the highest rates of exclusion errors at early deadlines if absolute familiarity is driving responses. If the hit rate advantage of LF words

over HF words is due to an increased influence of recollection for these items (as suggested by Joordens & Hockley, 2000; Reder et al., 2000) and the fast response deadline minimizes the recollective contribution, one might then expect an LF hit rate advantage only at the slower deadlines (cf. Balota et al., 2002).

In contrast to the above prediction, if participants also rely on a fast-acting *relative* familiarity process, then, at fast deadlines, one would expect a reversal of word frequency effects (i.e., *more* exclusion errors to LF than to HF words), because the relative boost in familiarity due to list exposure will be greater for LF than for HF to-be-excluded words. If both sources of familiarity contribute to exclusion errors, however, both frequency and repetition should exert an effect, with more errors to LF words and thrice-presented words. At the slower delays, recollection should be dominant, and so one should observe *fewer* exclusion errors to LF than to HF words. It is this crossover influence of word frequency across the deadlines that is critical for demonstrating a fast-acting relative familiarity process. Such findings would suggest that in addition to the baseline familiarity and recollection components proposed by the dual-process perspectives, there may be a third process that involves the assessment of a fast-acting relative boost in familiarity.

## Experiment 1

The first experiment employed a rhythmic-cuing paradigm that uses a rhythmic procedure to tightly constrain when responses are executed. Participants were cued to respond 500, 800, 1,000, or 1,500 ms after item presentation. These four timing conditions were selected to span the time course of episodic recognition decisions.

## Method

**Participants** The participants were undergraduate students enrolled in psychology courses at Washington University in

Saint Louis, who were compensated with course credit. All were native English speakers and had normal or corrected-to-normal vision. Originally, 65 participants were tested in the 500-ms and 800-ms conditions, 61 in the 1,000-ms condition, and 67 in the 1,500-ms condition. Because of high error rates and/or not attending to the rhythmic-cuing procedure, participants were eliminated from each timing condition, such that 60 participants remained in each, for a total of 240 subjects.

**Materials** A total of 90 HF and 90 LF words were selected from the Kučera and Francis (1967) norms. The words at each level of frequency were divided into six lists of 15 items, with the lists matched on concreteness, frequency, and length. However, inspection of the items after data collection was complete, using a better word frequency estimate (HAL; Burgess & Livesay, 1998), revealed that the frequency distributions overlapped. Hence, we excluded the 18 highest-frequency LF words and the 18 lowest-frequency HF words from the analyses, along with an additional 6 LF and 6 HF words to equate the lists on word length. Thus, the analyses were based on 11 items per list, instead of the 15 included in the study lists (see Table 2 for lexical characteristics of the stimuli). The same pattern of results was obtained when the full stimulus set was analyzed.

Two lists of HF items and two lists of LF items were assigned to the auditory condition, for a total of 30 HF and 30 LF items. In the visual condition, one HF and one LF list were studied once and one HF and one LF list were studied three times. The remaining two lists at each level of frequency were used as foils on the test. The six lists were rotated across experimental conditions and participants, such that each item appeared in each condition an equal number of times. An additional eight items at each frequency level were used as buffers.

Sixteen medium-frequency items were selected for a practice study and test phase to familiarize participants with the rhythmic-cuing paradigm. Eight items were studied, and all 16 were included in the practice recognition test.

**Table 2** Stimulus characteristics for the 11-item sets used in all experiments

Variable	Frequency		Difference (H – L)	<i>p</i> -value
	High	Low		
Length	6.03	6.32	–.29	.12
HAL log frequency	11.03	6.70	4.33	<.001
LDT RT (ELP)	607 ms	750 ms	–143 ms	<.001
LDT accuracy (ELP)	.98	.83	.15	<.001
Age of acquisition (MRC)	325	516	–191	<.001
Concreteness (MRC)	452	459	–7	.74
Familiarity (MRC)	575	389	186	<.001
Imageability (MRC)	496	470	26	.153

ELP: Values reported were obtained from the English Lexicon Project (Balota et al., 2007) database.

MRC: Values reported were obtained from the MRC database (Wilson, 1988).

**Procedure** Participants were tested individually. For each of the three phases (visual study, auditory study, test), instructions were presented on the computer screen while the experimenter read them aloud and the participant followed.

Participants were first given a practice study list and a recognition test in the appropriate timing condition. They were told that the experimental test would be similar and that the practice was to familiarize them with the procedure.

Practice items were presented visually one at a time for 1,500 ms, with an interstimulus interval of 500 ms. After the last item, participants were given test instructions specifying that, on each trial, a series of visual cues and auditory tones would be presented at a particular rhythm and that they should try to respond in rhythm. On each test trial, the following sequence of events occurred (see Fig. 1): (a) a blank screen appeared for 500 ms; (b) a pair of flankers (> and <) separated by approximately 14 spaces appeared in the center of the screen in synchrony with a 50-ms tone; (c) after the targeted interval (i.e., 500, 800, 1,000, or 1,500 ms), a second set of flankers appeared next to the first set (e.g., >> and <<) with a second tone; (d) a third set of flankers appeared with the third tone after the targeted interval; (e) a fourth set of flankers and a fourth tone were presented; (f) the test item appeared between the flankers concurrently with the last tone; (g) the flankers disappeared (the test item remained visible) and participants responded by pressing one of two keys on the keyboard (the “a” key for “old” responses and the “l” key for “new” responses); finally, (h) timing and accuracy feedback were presented for 1,000 ms before the next trial began. In the 1,500-ms condition, only three cuing tones were presented, in order to keep the total time of the test comparable to the other conditions. Pilot testing revealed that three tones were sufficient to allow participants to titrate their timing accurately. If the response was incorrect, the word WRONG appeared in the center of the screen. If participants correctly identified the item as old or new, an “O” appeared in the center of the screen. Feedback on timing was given in 50-ms increments. If a participant anticipated the deadline by 50 ms, a single minus sign appeared to the left of the “O” (“-O”). If participants were too slow by 100 ms, two plus signs

appeared to the right of the “O” (“O++”). A maximum of four plusses or minuses appeared. Participants were encouraged to use the feedback to improve their timing and accuracy. After the practice test, the experimental phase began.

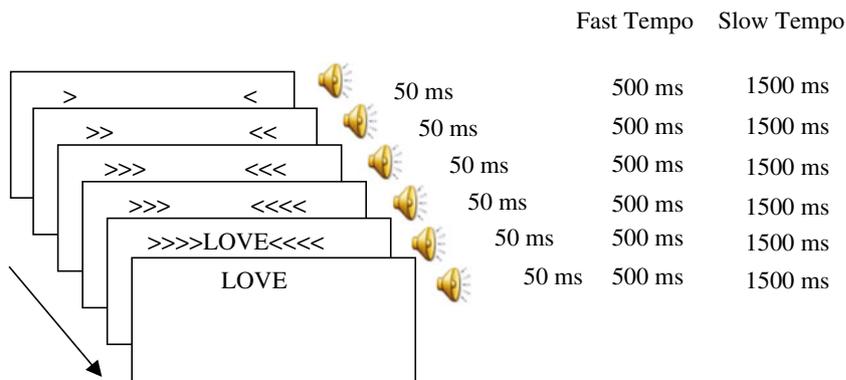
The visual study phase consisted of 120 items (half HF and half LF) and eight primacy and recency buffers. Participants read each item out loud. The presentation rate was the same as in the practice phase. The thrice-presented items in the visual list were equally distributed across each third of the list. Following the visual study list, the auditory study list began. Participants were instructed to repeat each word aloud and were informed that none of the words they heard were included in the visual study list. The auditory lists were digitally recorded in a female voice and spliced such that presentation occurred at a rate of one word every 2 s. Four primacy and recency buffers (two HF and two LF) were included. Visual and auditory study lists were presented in a new random order for each participant.

Immediately following the auditory list, test instructions appeared. Participants were told that they were to respond “old” only to the items they had heard. A “new” response was to be given to all other items, the visually studied items and any new items. Participants were told that if they remembered seeing the word, they could be sure they had not heard it and should therefore reject it. They were reminded to respond using the rhythmic cuing. The first few test trials included the primacy and recency items and some new foils matched to the buffers. The 180 critical items (60 visual, 60 auditory, and 60 new) were presented in random order. As in the practice phase, timing and accuracy feedback were given. Once instructions for the test had been given, the experimenter left the room. The entire session lasted approximately 45 min.

**Results and discussion**

The average RT deviation from the targeted response timing across participants was quite small. At the 500-, 800-, 1,000-, and 1,500-ms deadlines, the responses were

**Fig. 1** Sequence of events and timing in the rhythmic-cuing paradigm in Experiment 1



**Table 3** Average proportion “old” responses (*SD* in parentheses) as a function of timing condition, study condition, and frequency in Experiment 1

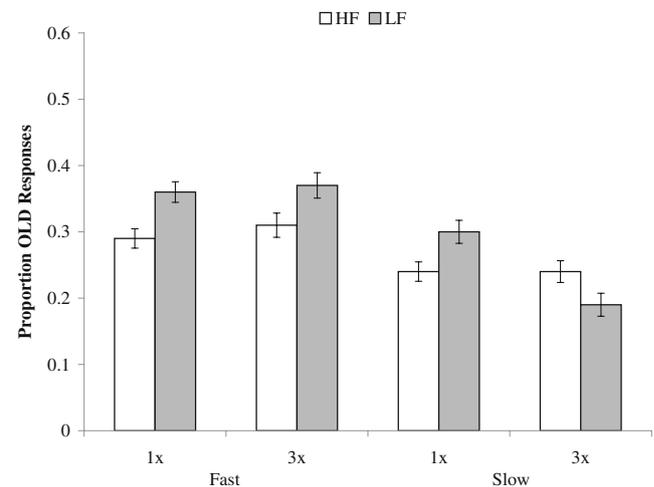
Timing	Study Condition and Frequency							
	Auditory		New		Visual 1x		Visual 3x	
	(Hits)		(False Alarms)		(Exclusion Errors)		(Exclusion Errors)	
	High	Low	High	Low	High	Low	High	Low
500	.30 (.14)	.36 (.15)	.28 (.15)	.25 (.16)	.29 (.16)	.38 (.17)	.31 (.18)	.38 (.19)
800	.31 (.15)	.48 (.17)	.21 (.15)	.22 (.20)	.29 (.16)	.35 (.18)	.31 (.21)	.36 (.22)
1,000	.30 (.15)	.47 (.19)	.18 (.14)	.17 (.17)	.25 (.16)	.31 (.18)	.24 (.16)	.23 (.20)
1,500	.35 (.16)	.55 (.17)	.18 (.13)	.15 (.15)	.24 (.15)	.29 (.18)	.24 (.19)	.17 (.14)

+36 ms, -8 ms, -14 ms, and -21 ms, respectively. Thus, we conclude that the rhythmic-cuing paradigm was very effective in controlling participants’ responses. The average proportions of “old” responses are displayed in Table 3.

In the following sections, we report analyses on accuracy data first, followed by analyses on measures of discriminability. Effects are significant at the .05 level unless otherwise noted. The proportions of “old” responses were submitted to a 2 (word frequency: HF vs. LF) x 4 (item type: hit, once-presented, thrice-presented, new) x 4 (timing condition: 500, 800, 1,000, 1,500 ms) mixed factor ANOVA. Other than the frequency x timing condition interaction, all effects were significant:  $F(1, 236) = 65.51, MSE = .03$ , for the effect of frequency;  $F(3, 708) = 99.67, MSE = .03$ , for the effect of item type;  $F(3, 236) = 4.1, MSE = .08$ , for the effect of timing condition;  $F(9, 708) = 10.84, MSE = .03$ , for the item type x timing condition interaction;  $F(3, 708) = 38.65, MSE = .02$ , for the frequency x item type interaction; and  $F(9, 708) = 4.8, MSE = .02$ , for the three-way interaction. The critical finding, of course, is the three-way interaction, which we examine in more detail below.

First, we examined the critical to-be-excluded items, which allow isolation of familiarity and recollection processes. To examine whether there were substantial differences between the fast (500 and 800 ms) and slow (1,000 and 1,500 ms) timing conditions, we conducted two separate ANOVAs with frequency, repetition, and timing condition as factors on the data from the two fast timing conditions (500 and 800 ms) and the two slow timing conditions (1,000 and 1,500 ms). There was no effect of timing condition, nor were there any interactions between timing condition and the other factors in these analyses (all  $F_s < 1.8$ , all  $p_s > .19$ ). Thus, after combining the data from the two fast and the two slow timing conditions, we conducted a 2 (frequency) x 2 (repetition) x 2 (timing condition: fast vs. slow) ANOVA on proportion of exclusion errors. All main effects and interactions were significant, all  $F_s > 4.5$ . We focus here on the critical three-way interaction between frequency, repetition, and timing condition,  $F(1, 238) = 4.7, MSE = .02$ .

As can be seen in Fig. 2, there are two points to note. First, at the fast timing conditions, there is a clear effect of word frequency,  $F(1, 118) = 21.26, MSE = .02$ , such that LF words ( $M = .37$ ) produce higher exclusion errors than HF words ( $M = .30$ ), suggesting an influence of relative familiarity at the fastest deadlines. If indeed one were relying on absolute familiarity, one would have expected exclusion errors to be greater for HF than for LF words. Interestingly, there is little influence of repetition, and no interaction, at these fast timing conditions, both  $F_s < 1.00$ . Second, at the slower timing conditions, where recollection should be engaged, a clear interaction between word frequency and repetition emerges. Specifically, repetition now decreases the exclusion errors, and this effect is primarily localized for the LF words, reflected by a reliable interaction between frequency and repetition,  $F(1, 118) = 14.34, MSE = .02$ . The interaction indicates that, at the slower timing conditions, there is a large effect of repetition for the LF words,  $t(119) = 5.2$ , but no effect for HF words,  $t < 1$ .



**Fig. 2** Average proportions of exclusion errors as a function of timing condition, frequency, and repetition in Experiment 1

To summarize, LF words produced more exclusion errors than HF words at the fast timing conditions, wherein familiarity-based processing should be engaged. This appears to reflect the influence of a relative change in familiarity, because absolute changes would have predicted the opposite pattern. In contrast, at the slow timing conditions, where recollection should be engaged, there was the expected frequency x repetition interaction, wherein repetition had a large influence in reducing the exclusion errors for LF words but not for HF words.

Turning to the hit rates, across all timing conditions these were higher for LF than for HF words (see Table 3),  $F(1, 236) = 194$ ,  $MSE = .01$ , and the size of this effect increased over the time course of recognition,  $F(3, 236) = 7.9$ ,  $MSE = .04$ . Specifically, LF hit rates increased over the time course from .36 to .55, whereas the HF hit rate only increased from .30 to .35,  $F(3, 236) = 8.8$ ,  $MSE = .01$ . The increase in hit rates for LF words across delays is consistent with the notion that recollection processes are more likely to be engaged at slower delays. Importantly, however, even at the fastest delay, there was a reliable hit rate advantage for LF words, supporting the notion that participants were relying on relative familiarity.

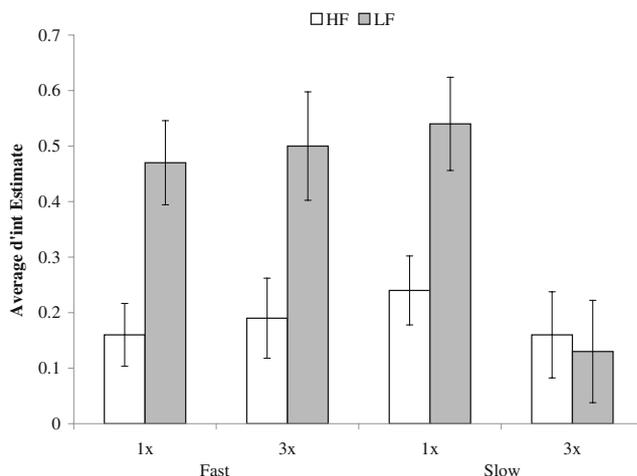
We also examined how reliance on absolute/baseline familiarity changed as a function of frequency and timing condition by examining the false alarms to foils. We expected an overall decrease in false alarms to foils as timing increased. In addition, we expected more false alarms to HF words than to LF words due to baseline familiarity differences. In fact, false alarms did decrease as a function of timing condition,  $F(3, 236) = 6.6$ ,  $MSE = .004$ , but the effect of frequency did not reach significance,  $F(1, 236) = 2.3$ ,  $p = .13$ , although false alarms to HF foils ( $M = .21$ ) were numerically higher than false alarms to LF foils ( $M = .19$ ). Although the lack of a reliable frequency effect for the foils was unexpected, and will be further addressed in the subsequent experiments, it is noteworthy that Hintzman, Caulton, and Curran (1994) found that the false alarm difference as a function of word frequency did decrease when a speeded response was imposed (see their Fig. 2, p. 281). It is possible that the strong attentional demands of the rhythmic-cuing paradigm used in Experiment 1 may have minimized the word frequency effect for nonstudied foils.

We also computed estimates of  $d'$  to examine changes in discriminability. Following McElree et al. (1999), we computed a  $d'_{\text{int}}$  estimate, in which we scaled exclusion errors against false alarms. The  $d'_{\text{int}}$  scaling provides a measure of the degree to which exclusion errors exceed false alarm rates to foils. In this case, higher  $d'_{\text{int}}$  estimates reflect worse performance, because they indicate that the exclusion errors exceed the baseline false alarm rate. As in the analyses on raw scores, we collapsed across the two fast and the two slow timing conditions after initial analyses

revealed no effects of timing condition and no interactions, all  $F_s < 2.4$ , all  $p_s > .12$ .

The  $d'_{\text{int}}$  estimates were submitted to a 2 (frequency) x 2 (item type: once-presented, thrice-presented) x 2 (timing condition: fast vs. slow) mixed factor ANOVA. As in the analyses on raw scores, all effects other than the frequency x timing condition interaction were reliable, all  $F_s > 5.8$ . Here, we focus on the three-way interaction,  $F(2, 238) = 5.7$ ,  $MSE = .28$ . As shown in Fig. 3, this pattern is consistent with the proportion “old” analyses. Specifically, at the fast timing conditions, LF words resulted in higher  $d'_{\text{int}}$  estimates than HF words,  $F(1, 119) = 19.19$ , and there is no evidence of an effect of repetition,  $F < 1$ , and no frequency x repetition interaction,  $F < 1$ . However, at the slower deadlines, the effect of repetition is reliable,  $F(1, 119) = 18.2$ , and frequency interacts with repetition,  $F(1, 119) = 11.8$ . As noted above, higher estimates reflect worse performance relative to new items, consistent with our hypothesis that the relative familiarity changes for LF words, as compared to HF words, are lowering performance at the fast deadlines, whereas at the slow deadlines, recollection processes that benefit thrice-presented LF words are engaged.

In sum, at fast deadlines, LF words produced higher exclusion errors (and hit rates) than HF words, suggesting a fast-acting relative familiarity process, whereas at the slow response deadlines, there were effects of both frequency and repetition with clear interactive effects, supporting a recollection-based process. Thus, it appears that both relative familiarity and recollection are playing a role in this two-list exclusion task. Further supporting the contribution of relative familiarity, we observed an LF advantage in hit rates, even at the fastest timing conditions, that increased once recollection was also contributing. Turning to the new items, although the means were in the predicted



**Fig. 3**  $d'_{\text{int}}$  estimates as a function of timing condition, frequency, and repetition in Experiment 1

direction, the effect of word frequency did not reach significance, thereby putting in question the role of baseline familiarity and suggesting that participants were relying primarily on assessments of relative familiarity (which do not vary for new items). We further explore this finding in the following experiments.

**Experiment 2**

As noted, the attentional demands of the strict rhythmic-cuing procedure used in Experiment 1 might have minimized the influence of word frequency on the lure items and overall have decreased reliance on absolute/baseline familiarity (cf. the hit rate advantage at all timing conditions and the lack of a repetition effect at fast timing conditions in Experiment 1). In Experiment 2, we eliminated the rhythmic-cuing procedure and used a more standard deadlining procedure, which should impose fewer demands on the attentional system (because participants do not have to maintain the rhythmic timing). In this experiment, we explored a fast 800-ms and a slow 1,500-ms deadline, because previous research (e.g., Jacoby et al., 1998) has shown that with these deadlines one can find evidence of familiarity-based processing at the fast deadline and recollection-based processing at the slow deadline, and because the two fast and two slow timing conditions in Experiment 1 did not differ from one another.

**Method**

*Participants* Ninety-six undergraduate students (48 in each deadline condition) from the same population participated for course credit.

*Materials* The same materials were used, with the exception of one word (*flower* was used instead of *chair* as an HF item).

*Procedure* The same general procedure was used, with the following exceptions. Repetition of items in the visual list ranged from 10 to 35 items later. In the test, equal numbers of heard, read, and new items were presented in each third

of the test list, with no more than three items of the same type presented in a row. With these constraints, a pseudorandom order was created for both the study and test lists, and the same order was presented to all participants. The auditory lists were recorded in a female voice onto an audiocassette, at a rate of one word every 2 s.

Participants in the slow deadline condition were instructed to wait for the appearance of a row of asterisks before making their response and were told that responses made before the asterisks appeared were invalid, as were responses occurring more than 800 ms after the asterisks. In the latter case, a warning tone sounded and the message “Too slow” appeared on the screen. No accuracy feedback was given.

Participants in the fast deadline condition were instructed to respond as quickly as possible as soon as the word appeared on the screen. Responses slower than 800 ms resulted in a warning tone and a “Too slow” message. In both conditions, participants were told to avoid receiving the warning message and tone and to guess “old” if unsure.

**Results and discussion**

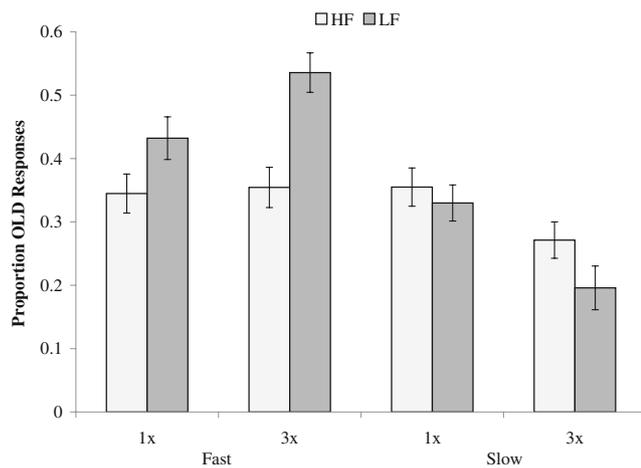
Only responses made within the 800- and 1,500-ms response windows were included in the analyses. Overall, 10% and 3% of the data points were excluded in the fast and slow conditions, respectively. The remaining data were trimmed by omitting all responses that exceeded each participant’s mean response latency by 2.5 SDs (1.4% and 2.5% in the fast and slow conditions, respectively).

Table 4 displays the average proportions of “old” responses. “Old” responses were submitted to a 2 (frequency) x 2 (deadline) x 4 (item type) mixed ANOVA. All main effects and interactions were significant, all *F*s > 8.8, and hence we focus on the theoretically driven comparisons here.

First, consider the critical exclusion errors. As shown in Fig. 4, at the fast deadline, there were more exclusion errors to thrice-presented LF words than to once- or thrice-presented HF words and to once-presented LF words. As indicated by the frequency x repetition interaction, *F*(1, 47) = 13.8, there is no influence of repetition for the HF words at

**Table 4** Proportion “old” responses (*SD* in parentheses) as a function of timing, study condition, and frequency in Experiment 2

Timing	Study Condition and Frequency							
	Auditory		New		Visual 1x		Visual 3x	
	(Hits)		(False Alarms)		(Exclusion Errors)		(Exclusion Errors)	
	High	Low	High	Low	High	Low	High	Low
Fast	.42 (.21)	.55 (.21)	.22 (.16)	.21 (.16)	.39 (.22)	.43 (.22)	.38 (.22)	.56 (.21)
Slow	.56 (.20)	.76 (.15)	.29 (.21)	.22 (.22)	.36 (.22)	.33 (.21)	.27 (.21)	.20 (.24)



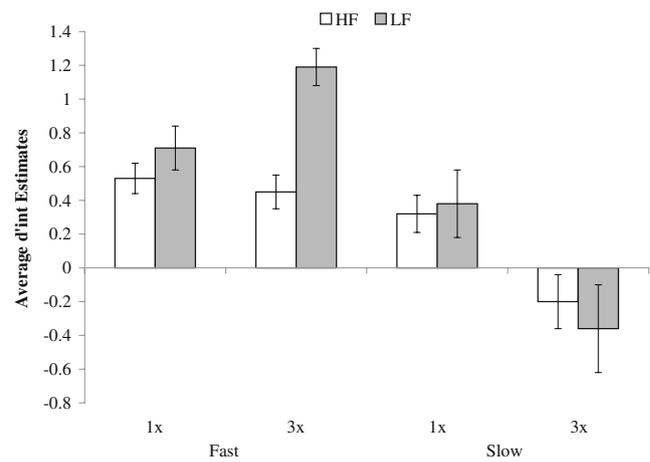
**Fig. 4** Average proportions of exclusion errors as a function of timing condition, frequency, and repetition in [Experiment 2](#)

the fast deadline,  $t < 1$ , but a large increase in errors for thrice-presented LF words,  $t(47) = 4.6$ .

Turning to the slow deadline, a very different pattern emerges: Main effects of frequency and repetition,  $F_s > 6.0$ , but no interaction,  $F < 1$ . Specifically, errors to both HF and LF words declined as a function of repetition,  $t(47) = 2.92$  and  $4.77$ , respectively. In sum, there are large and opposite effects of repetition for the more familiarity-driven fast deadline and the more recollection-based slower deadline conditions. Thus, with a less-demanding deadlining procedure, the effect of repetition on exclusion errors did reemerge, replicating the findings of Jacoby et al. (1998). Obtaining the repetition effect for HF words at the slow deadline does lend some support to the notion that, in [Experiment 1](#), participants were relying primarily on frequency-modulated changes in relative familiarity (which, as expected, primarily affected LF words).

Turning to the hit rates, at both response deadlines LF hits exceeded HF hits,  $F(1, 94) = 94.1$ ,  $p < .001$ , although, as expected, the difference was more pronounced at the slower than at the faster deadline,  $F(1, 94) = 5.56$ ,  $p = .02$ . Hence, consistent with [Experiment 1](#), the typical LF advantage was found for hit rates, and this advantage increased across delays as the contribution of recollection increased.

Turning to the lure data, we now obtained reliably more false alarms to HF foils ( $M = .26$ ) than to LF foils ( $M = .21$ ),  $F(1, 94) = 5.02$ ,  $p = .03$ , although overall false alarm rates did not differ as a function of deadline,  $F = 1.5$ . The interaction between frequency and deadline was not significant,  $F(1, 94) = 2.1$ ,  $p = .15$ , but it was in the predicted direction (i.e., a larger frequency effect at the slow deadline than at the fast deadline). In sum, under the more lenient deadlining procedure, we now obtained the LF advantage for correct rejections, suggesting that under these conditions absolute/baseline familiarity exerts an influence.



**Fig. 5**  $d'_{int}$  estimates as a function of timing condition, frequency, and repetition in [Experiment 2](#)

Turning to the  $d'_{int}$  scaling, the estimates were submitted to a 2 (HF vs. LF)  $\times$  2 (once-presented vs. thrice-presented)  $\times$  2 (fast vs. slow deadline) ANOVA. Once again, the critical three-way interaction was reliable,  $F(1, 94) = 8.79$ , so, for the sake of brevity, we focus on its interpretation. As can be seen in [Fig. 5](#), at the fast deadline, LF words yielded higher values than HF words,  $F(1, 47) = 13.18$ , and a much larger repetition effect,  $F(1, 47) = 14.10$ , consistent with the fact that the thrice-presented LF words were falsely recognized at high rates and that increases in relative familiarity were driving the effect. At the slow deadline, however, only the effect of repetition was reliable:  $d'_{int}$  values for once-presented items reflected worse discriminability than new items,  $F(1, 47) = 34.91$ , but thrice-presented items were better recognized, regardless of frequency. These results indicate that, at the slow deadline, repetition seems to be influencing recognition more than does frequency, given that both HF and LF words yielded the same patterns of results.

In sum, [Experiment 2](#) suggests that the rhythmic-cuing paradigm employed in [Experiment 1](#) maximized reliance on relative familiarity assessments, at some cost to assessments of absolute/baseline familiarity. When the strict rhythmic-cuing paradigm was removed, the repetition  $\times$  frequency interaction in exclusion errors emerged at the fast deadlines. Moreover, there was a main effect of frequency on correct rejections, which was larger at the slow deadline, suggesting that participants did rely on baseline familiarity. Thus, the fact that participants produced more exclusion errors to LF than to HF words at the fast deadline, and at the same time produced more false alarms to new HF than to new LF words, suggests that both relative familiarity and baseline familiarity are indeed engaged in this task. In [Experiment 3](#), we examined performance in the two-list exclusion paradigm in the absence of timing manipulations, to obtain a

clearer picture of how frequency and repetition interact in this paradigm.

### Experiment 3

In this final experiment, we further explored the role of the rhythmic-cuing paradigm on assessments of absolute/baseline familiarity. As suggested above, and as indicated by the results of Experiment 2, the cuing paradigm (combined with the demands of the exclusion instructions) appears to have driven participants to decrease reliance on absolute/baseline familiarity, thus eliminating the repetition effect at fast deadlines and the false alarm portion of the word frequency mirror effect in Experiment 1. In Experiment 2, we used a less strict deadlining procedure and found effects of both repetition and frequency at the fast deadline, along with the predicted effect of word frequency on lure items. If the cuing paradigm modulates to some degree the reliance on baseline familiarity, one might then expect the false alarm portion of the mirror effect to increase even further when deadlines are completely removed. This was the goal of Experiment 3.

#### Method

**Participants** Thirty-six participants from the same population were recruited in exchange for partial course credit.

**Materials and procedure** The materials and procedure were the same as in Experiment 1, except that there was no manipulation of response deadline (participants were encouraged to respond quickly, but not at the expense of accuracy), although accuracy feedback was still provided. The practice phase was included; the test instructions were the same (i.e., participants were told to reject the visually presented items).

#### Results

The average response latency, across all response and item types, was 893 ms. The proportions of “old” responses are presented in Table 5.

An ANOVA with Frequency and Item Type as factors was conducted. The main effect of frequency was not reliable,  $F(1, 35) = 2.15$ ,  $p = .15$ , but the effect of item type was,  $F(3, 105) = 42.58$ , and, most importantly, the interaction was highly reliable,  $F(3, 105) = 19.59$ . Hit rates and false alarm rates to new foils showed the expected mirror effect: higher hits and lower false alarm rates to LF than to HF words,  $t(35) = 7.3$  and 2.3, respectively.

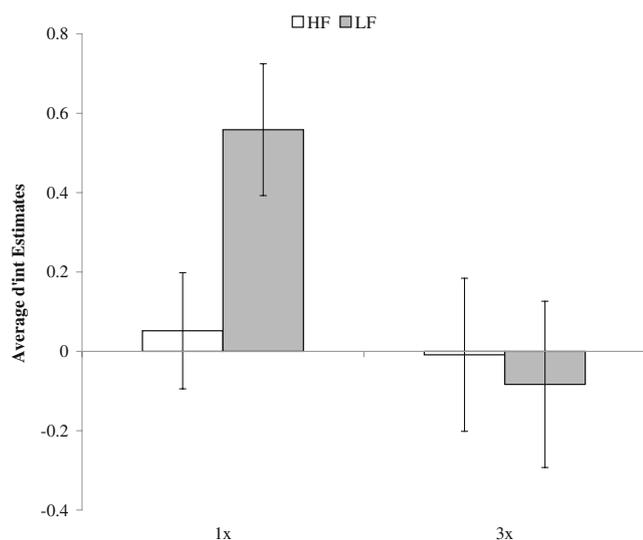
In a separate ANOVA including only the exclusion errors, the effect of frequency was not significant,  $F < 1$ , but the effect of repetition was,  $F(1, 35) = 4.4$ , reflecting more errors for once-presented ( $M = .25$ ) than for thrice-presented words ( $M = .21$ ). The interaction was also significant,  $F(1, 35) = 8.2$ . There were slightly more errors to once-presented LF words than to once-presented HF words,  $t(35) = 1.1$ ,  $p = .26$  (although the difference was not reliable, it was in a direction consistent with relative familiarity increases in the absence of sufficient information for a recollection-based rejection), but significantly fewer errors to thrice-presented LF words than to thrice-presented HF words,  $t(35) = 2.3$ . As in Experiment 1, repetition had no effect on HF exclusion errors but a reliable and positive effect on LF errors.

The emergence of the false alarm portion of the mirror effect, coupled with the reduced error rate—relative to Experiment 1—for once-presented LF words, appears to be consistent with the notion that the cuing paradigm was engaging additional attentional demands. When participants were not instructed to respond in time with the cues, they could rely on baseline familiarity to reject new items, thus resulting in the LF benefit in correct rejections.

The  $d'_{\text{int}}$  scaling (see Fig. 6) revealed no effect of frequency,  $F < 1.6$ ; a significant effect of repetition,  $F(1, 35) = 10.1$ , such that once-presented words showed worse discriminability than thrice-presented words; and the interaction between frequency and repetition observed at the slow deadlines in Experiment 1,  $F(1, 35) = 8.77$ : For once-presented foils, discriminability was worse for LF than for HF words,  $t(35) = 2.7$ , but for thrice-presented words, there was no effect of frequency,  $t < 1$ . Another way of interpreting this pattern is that only LF words showed a benefit of repetition,  $t(35) = 4.1$ ; for once-presented items, changes in relative familiarity drove up errors, but for thrice-presented items, recollection could counteract the increase in familiarity. This finding also replicated the

**Table 5** Proportion “old” responses (*SD* in parentheses) as a function of study condition and frequency in Experiment 3

Frequency	Study Condition			
	Auditory (Hits)	New (False Alarms)	Visual 1x (Exclusion Errors)	Visual 3x (Exclusion Errors)
High	.37 (.14)	.22 (.16)	.23 (.13)	.24 (.18)
Low	.57 (.18)	.15 (.17)	.26 (.14)	.17 (.16)



**Fig. 6**  $d'_{int}$  estimates as a function of frequency and repetition in Experiment 3

effect observed in Experiment 1, that HF words underwent little change in relative familiarity and produced relatively little benefit of recollection processes, as compared to LF words. Interestingly, HF exclusion errors did not differ from false alarms to new HF foils at either level of repetition, both  $t_s < 1$ , confirming that these items do not benefit from recollection, nor do they undergo substantial changes in familiarity, regardless of the number of presentations.

In sum, Experiment 3 yielded two main findings. First, under a no-deadline condition, we obtained a large word frequency mirror effect in hits and false alarms, consistent with our suggestion that assessments of absolute/baseline familiarity were reduced in Experiment 1 due to the attentional demands of the rhythmic-cuing procedure. Second, these results replicate the overall pattern of results at the slow delays in Experiment 1, yielding repetition effects only for LF words, supporting the notion that such items are both better recollected *and* undergo larger changes in relative familiarity.

### General discussion

The present series of experiments was designed to examine the possibility of distinct types of familiarity and their time courses in episodic recognition. Dual-process models of recognition assume that recollection and familiarity are both involved in recognition decisions. The word frequency effect has been a test case for many models of recognition, and in recent dual-process models that have been proposed to account for the word frequency mirror effect (Joordens & Hockley, 2000; Reder et al., 2000), HF words result in increased false alarm rates because of higher baseline familiarity, whereas LF words elicit more hits because they

are easier to recollect, due to fewer preexperimental contexts and/or fewer associations. The focus in these accounts, therefore, has been on a fast-acting baseline familiarity process and a slower, more resource-demanding recollective process.

However, most past studies have used paradigms in which familiarity and recollection act in concert (e.g., Joordens & Hockley, 2000; Reder et al., 2000). Moreover, in previous studies in which the two processes were placed in opposition (e.g., Guttentag & Carroll, 1994), the time courses of the effects were not investigated, nor were manipulations aimed at systematically varying different sources of familiarity. In the present studies, we addressed both of these issues by using the two-list exclusion paradigm (Jacoby, 1999) and a response deadlining procedure to examine the effect of familiarity when it is placed in opposition to recollection across the time course of recognition decisions. In addition to varying preexperimental familiarity (i.e., word frequency), we manipulated experimental familiarity by presenting each word either once or three times, thus allowing us to examine relative changes in familiarity as well as absolute/baseline levels.

The exclusion error results in the present study were most telling. These errors reflect situations in which familiarity is operating when recollection has failed. In Experiments 1 and 2, more errors were made at the fast deadlines to LF than to HF words. Because LF words undergo a larger increase in *relative* familiarity due to a study event, compared to HF words (which are already near asymptotic familiarity levels), the effect of frequency provides strong evidence for the role of a fast-acting relative familiarity process. Another way of conceptualizing this is that a study event can temporarily increase the familiarity signal of LF words such that they actually produce a stronger signal than HF words, in the sense that these items undergo a larger change from their baseline state. In Experiment 2, we were able to replicate earlier findings of an ironic effect of repetition on exclusion errors. That is, under a more lax deadline procedure, an effect of absolute/baseline familiarity reemerged under the fast deadline, reflecting a change in reliance on familiarity due to repetition (i.e., exclusion errors to thrice-presented LF words were higher than errors to once-presented LF words). Interestingly, the effect was limited to LF words, consistent with the notion that HF words are near an asymptotic level of activation and are less affected by recent presentation. Thus, the more commonly used deadline procedure in Experiment 2 allowed us to assess the contributions of relative and absolute familiarity separately, but once additional resources were tied up by the rhythmic-cuing paradigm used in Experiment 1, only the relatively automatic assessment of relative familiarity drove responses (and, under these conditions, no effects of repetition were

seen for HF words, and for LF words, the critical aspect seems to have been “studied/nonstudied” rather than “once/thrice presented”). At fast deadlines with the strict rhythmic-cuing procedure, it appears that LF words captured attention more than HF words, independent of whether the stimulus was repeated once or thrice. At the slow deadlines, however, repetition and word frequency both contributed to influence performance, precisely as one would expect from a more recollection-based process. Specifically, only LF items showed consistently reliable reductions in exclusion errors due to repetition, whereas HF items were relatively uninfluenced by repetitions.

In **Experiment 3**, we examined performance in the exclusion paradigm without time pressure to further address the effect of the additional cognitive load imposed by the timing paradigm. As expected, the results from this experiment generally replicated those from the slow deadlines in **Experiments 1** and **2**, with larger beneficial effects of repetition for LF than for HF words. Importantly, along with **Experiment 2**, this experiment yielded a larger influence of word frequency on the portion of the mirror effect from false alarms to new foils, indicating a reliance on baseline familiarity processes when there was no other diagnostic information available, as assumed by most dual-process models.

The main finding from the present experiments is that at least two distinct components of familiarity can contribute to the influence of word frequency in recognition memory. Absolute/baseline and relative familiarity exert dissociable effects on recognition decisions. Of course, the notion of a relative familiarity change as a useful metric for recognition decisions is not new. For example, Brown, Lewis, and Monk (1977) originally proposed that changes in relative familiarity account for the word frequency effect in recognition memory. Jacoby and Dallas (1981); (see also Whittlesea, 2002; Whittlesea & Williams, 2001) argued that the discrepancy between the perceived and expected fluency of processing can be used as a signal for “oldness.” Indeed, Jacoby and Dallas demonstrated that the influence of relative fluency was particularly strong for LF items in both perceptual identification and episodic recognition, suggesting that the change in relative processing fluency served as a basis for recognition for this class of items. Likewise, Benjamin, Bjork, and Hirshman (1998) argued that the difference between perceived fluency and expected fluency can be a signal for a recent exposure to a stimulus. The change in relative familiarity for LF compared to HF words might be considered analogous to the perceived-versus-expected-fluency idea. Balota et al. (2002) proposed a similar mechanism in accounting for the changes in memory performance in older adults and in early-stage Alzheimer’s disease. Although these perspectives have not been specific regarding the role of attention in the

computation of relative changes in familiarity, one might have expected such a computation to demand more controlled processing. In this light, an important contribution of the present work is to demonstrate that such relative familiarity changes come on early in processing. Hence, the computation of the relative change in perceived versus expected familiarity does not need to be a volitional, resource-demanding process. Indeed, the fact that relative familiarity effects are apparent very early in the time course of recognition in both hits and exclusion errors suggests that such a process is relatively automatic. In the present work, the distinction we emphasize is between preexperimental and experimentally induced changes in familiarity, and at least the results from one experiment provide tentative evidence for such a dissociation.<sup>1</sup> Clearly, further work will need to substantiate this finding, but most importantly, this does not diminish the importance of the observation of a fast-acting relative familiarity effect.

The dual-process models outlined in the introduction attribute the correct-rejection advantage for LF words to lower baseline familiarity levels and the hit rate advantage to recollection. To our knowledge, neither model has a relative familiarity process. It is possible that the SAC model (Reder et al., 2000) could accommodate the present results by assuming that accessing the event node, which supports the recollection advantage for LF words, takes time, but at fast deadlines the familiarity signal from the word node drives responses (although, from their Fig. 2, it appears that the word node for HF nodes is stronger, at least until 10 presentations have occurred). Our proposed relative familiarity process, rather than being subsumed under the umbrella term *familiarity*, as in earlier models, is a distinct process that influences both hits and errors when source discriminations are required but not available. As noted in the introduction, standard recognition paradigms do not allow one to separate the distinct influences of multiple processes, whereas the exclusion paradigm does.

If indeed baseline and relative familiarity are distinct sources of information, it is possible that individuals can modulate the extent to which they rely on one or the other. Furthermore, switching between assessments of absolute and relative familiarity might require additional resources. Where the effects of absolute familiarity should be evident is in false alarm rates to totally new items. As noted, in **Experiment 1** we found only a marginal effect of word frequency at the slowest deadline. The absence of an LF advantage in the false alarm portion of the mirror effect is

<sup>1</sup> An alternative possibility is that absolute/baseline familiarity is available more quickly but needs to be assessed against some criterion, whereas relative familiarity might require more time to initially compute, but that computation already involves some assessment, so demands are less for this second operation. We thank William Hockley for suggesting this alternative.

unlikely to be due to insufficient power. Using the G\*Power package (Erdfelder, Faul, & Buchner, 1996), the observed power in the analysis on false alarms to new foils was .88 to detect a small effect (Cohen's  $d$  of 0.20) at the .05 level. We have suggested that the rhythmic-cuing paradigm increased participants' reliance on relative familiarity. The higher rates of exclusion errors to LF than to HF items and the consistent hit rate advantage for LF items are consistent with this hypothesis. It appears that when attentional resources are tied up by a complex timing paradigm, participants responded primarily on the basis of relative familiarity, in which case no reliable difference between new HF and LF words would be expected. As noted above, in Experiment 2, with a less-demanding deadlining paradigm, and in Experiment 3, in which no deadline was imposed, the false alarm portion of the mirror effect reemerged. It seems that with the additional demands of the rhythmic-cuing paradigm, participants were primarily responding on the basis of the prepotent and obligatory assessment of relative familiarity, which appears to require few attentional resources. Indeed, the lack of a repetition effect at fast deadlines in Experiment 1 and the absence of a false alarm effect are consistent with this hypothesis. Once additional resources are made available, either by extending the deadline or reducing the cognitive load via the more lax deadlining procedure, all three processes (relative familiarity, absolute/baseline familiarity, and recollection) can exert an influence.

In summary, the present results indicate three distinct influences of word frequency on recognition memory performance. Current dual-process models emphasize the importance of a relatively fast baseline familiarity process (to account for the LF advantage in false alarms) and a relatively slower recollection-based process (to account for the LF advantage in hits). The present results indicate that participants are also influenced by a fast computation of a relative change in familiarity, which appears to require fewer resources than a computation of baseline familiarity. In the past, one might have expected such a process to be subsumed under recollection, which compares the expected familiarity with the current familiarity. However, the present results indicate that such a computation is very fast-acting and is quite distinct from the recollection-based process, in that it does not involve retrieval of source or context information and has a markedly distinct time course.

## References

- Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17*, 814–823.
- Atkinson, R. C., & Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Fourth International symposium on attention and performance* (pp. 583–611). New York: Academic.
- 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*. San Francisco: Freeman.
- Balota, D. A., Burgess, G. C., Cortese, M. J., & Adams, D. R. (2002). The word-frequency mirror effect 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.
- Balota, D. A., Cortese, M. J., Hutchison, K. A., Neely, J. H., Nelson, D., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project: A Web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords. *Behavior Research Methods, 39*, 445–459. Available at <http://ellexicon.wustl.edu/>
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*, 283–316.
- Balota, D. A., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 576–587.
- Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General, 128*, 32–55.
- Benjamin, A. S., Bjork, R. A., & Hirshman, E. (1998). Predicting the future and reconstructing the past: A Bayesian characterization of the utility of subjective fluency. *Acta Psychologica, 98*, 267–290.
- Benjamin, A. S., & Craik, F. I. M. (2001). Parallel effects of aging and time pressure on memory for source: Evidence from the spacing effect. *Memory & Cognition, 29*, 691–697.
- Boldini, A., Russo, R., & Avons, S. E. (2004). One process is not enough! A speed-accuracy tradeoff study of recognition memory. *Psychonomic Bulletin & Review, 11*, 353–361.
- Brown, J., Lewis, V. J., & Monk, A. F. (1977). Memorability, word frequency, and negative recognition. *The Quarterly Journal of Experimental Psychology, 29*, 461–473.
- Burgess, C., & Livesay, K. (1998). The effect of corpus size in predicting reaction time in a basic recognition task: Moving on from Kučera and Francis. *Behavior Research Methods, Instruments, & Computers, 30*, 272–277.
- Dewhurst, S. A., Holmes, S. J., Brandt, K. R., & Dean, G. M. (2006). Measuring the speed of the conscious components of recognition memory: Remembering is faster than knowing. *Consciousness and Cognition, 15*, 147–162.
- Diana, R. A., Reder, L. M., Arndt, J., & Park, H. (2006). Models of recognition: A review of arguments in favor of a dual-process account. *Psychonomic Bulletin & Review, 13*, 1–21.
- Erdfelder, E., Faul, F., & Buchner, A. (1996). G\*POWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers, 28*, 1–11.
- Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13*, 8–20.
- Guttentag, R. E., & Carroll, D. (1994). Identifying the basis for the word frequency effect in recognition memory. *Memory, 2*, 255–273.
- Hintzman, D. L., & Caulton, D. A. (1997). Recognition memory and modality judgments: A comparison of retrieval dynamics. *Journal of Memory and Language, 37*, 1–23.
- Hintzman, D. L., Caulton, D. A., & Curran, T. (1994). Retrieval constraints and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 275–289.
- 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. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 3–22.
- Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical and perceptual learning. *Journal of Experimental Psychology: General*, *110*, 306–340.
- Jacoby, L. L., Jones, T. C., & Dolan, P. O. (1998). Two effects of repetition: Support for a dual-process model of know judgments and exclusion errors. *Psychonomic Bulletin & Review*, *5*, 705–709.
- 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: Brown University.
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*, 252–271.
- 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.
- Reder, L. M., Nhuoyvanisvong, 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.
- Reder, L. M., Paynter, C., Diana, R. A., Ngiam, J., & Dickison, D. (2007). Experience is a double-edged sword: A computational model of the encoding/retrieval tradeoff with familiarity. In B. Ross & A. Benjamin (Eds.), *The psychology of learning and motivation* (Vol. 48, pp. 271–312). San Diego: Academic.
- Whittlesea, B. W. A. (2002). Two routes to remembering (and another to remembering not). *Journal of Experimental Psychology: General*, *131*, 325–348.
- Whittlesea, B. W. A., & Williams, L. D. (2001). The discrepancy-attribution hypothesis: II. Expectancy, uncertainty, surprise, and feelings of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 14–33.
- Wilson, M. (1988). The MRC Psycholinguistic Database: Machine-usable dictionary, Version 2.00. *Behavior Research Methods, Instruments, & Computers*, *20*, 6–11.
- Yonelinas, A. P. (2001). Consciousness, control, and confidence: The three Cs of recognition memory. *Journal of Experimental Psychology: General*, *130*, 361–379.
- 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. (1994). Dissociations of processes in recognition memory: Effects of interference and of response speed. *Canadian Journal of Experimental Psychology*, *48*, 516–535.
- Yonelinas, A. P., & Jacoby, L. L. (1996). Noncriterial recollection: Familiarity as automatic, irrelevant recollection. *Consciousness and Cognition*, *5*, 131–141.

#### Author Note

Portions of this research were presented at the 2006 Meeting of the Psychonomic Society in Houston, Texas. We thank Jason Arndt, William Hockley, Bruce Maxwell, and several anonymous reviewers for helpful comments on earlier versions of the manuscript.