

# Episodic feeling-of-knowing resolution derives from the quality of original encoding

CHRISTOPHER HERTZOG

*Georgia Institute of Technology, Atlanta, Georgia*

JOHN DUNLOSKY

*Kent State University, Kent, Ohio*

AND

STARLETTE M. SINCLAIR

*Georgia Institute of Technology, Atlanta, Georgia*

In recent studies, researchers have argued for adult age-related deficits in the resolution of episodic feeling of knowing (FOK) owing to a decline in inferential processes. In the present study, we introduce the memory constraint hypothesis, which argues that deficits are an outcome of differences in the level of learning. A repetition delay paradigm for a list of paired-associate items showed that repeated presentations at encoding increased memory performance and in turn increased FOK resolution for unrecalled items. Older adults who were given a 48-h delay between encoding and subsequent tests (and FOKs) had equivalent memory performance to younger adults who were given a 7-day delay. In this case, age equivalence arose in FOK resolution except at the lowest levels of recognition in the single-presentation condition. The use of effective strategies during encoding correlated with memory performance and FOKs, even for unrecalled pairs. These results are inconsistent with an inferential-deficit explanation of age deficits in FOK resolution; they point to the importance of original encoding quality as a potent contributor to FOK resolution, and they argue for equating age groups on memory performance when evaluating the episodic FOK resolution of age differences.

The feeling of knowing (FOK) is a state in which individuals have a subjective experience that they know information that they cannot presently recall. Following the results of Hart (1965), FOK has been measured by having people make FOK ratings after attempts to recall targets have failed. Often, FOKs are scaled as a percentage confidence in the likelihood of future recognition of the targets. Semantic FOKs typically follow attempts to recall world knowledge or facts (see, e.g., Butterfield, Nelson, & Peck, 1988). Episodic FOKs are often generated by tests of cued recall for paired-associate (PA) items (e.g., DOG-SPOON) in which a cue (DOG-?) is used to elicit both the recall attempt and the FOK. Thus, a major distinction between semantic and episodic FOKs is that the former are made for unrecalled information that was typically learned prior to the experiment, possibly through multiple exposures, and possibly with repeated subsequent access so that the information can be treated as knowledge stored in semantic memory. For episodic FOKs, the items being judged were first studied in the particular context of the experiment with the features of encoding, retention, and retrieval under experimental control. In the present study, we evaluated sources of influence on episodic FOKs for recently learned, verbal paired associates (e.g., Schacter, 1983).

FOKs are typically validated by correlating them with recognition memory for unrecalled items. Within a given individual, *resolution* is typically defined as the correlation of FOKs with recognition memory outcomes (e.g., successes vs. failures), and it indexes the degree to which higher FOKs are more likely for successfully recognized targets. It has become traditional in the metacognitive literature to operationally define resolution by average (mean) intraindividual Goodman-Kruskal gamma correlations (T. O. Nelson, 1984).<sup>1</sup>

## Aging and Episodic FOKs

Aging has little or no effect on the resolution of semantic FOKs (Allen-Burge & Storandt, 2000; Bäckman & Karlsson, 1985; Butterfield et al., 1988; Lachman, Lachman, & Thronesbury, 1979; Marquié & Huet, 2000; Souchay, Moulin, Clarys, Taconnat, & Isingrini, 2007). In recent years, however, a number of investigators have suggested that the resolution of episodic FOKs is impaired in old age (e.g., Perrotin, Tournelle, & Isingrini, 2008; Souchay, Isingrini, & Espagnet, 2000; Souchay et al., 2007). All of these studies involved a single study episode for stimuli with equivalent experimental conditions (list length, study procedures, test delay) that was given to young and old

adults. Under such conditions, age deficits in associative memory are typically observed (Kausler, 1994), as they were in the FOK studies just cited. The Souchay et al. (2007) study was particularly interesting concerning age-related impairments in episodic FOKs, because their older adults showed an advantage in semantic FOK resolution but a deficit in episodic FOK resolution.

The hypothesis of age-related deficits in episodic FOK accuracy is something of an exception in the area of aging and episodic metacognitive monitoring, in which age equivalence in the accuracy of basic metacognitive judgments such as FOKs, judgments of learning, or recognition confidence judgments (CJs) are often found despite impairments in memory itself (see, e.g., Hertzog, Kidder, Powell-Moman, & Dunlosky, 2002; Hertzog, Sinclair, & Dunlosky, 2010; Hines, Touron, & Hertzog, 2009). The episodic FOK deficit hypothesis has gained traction in part because of its plausible connection with evidence that (1) aging affects cognitive control mechanisms that are associated with frontal lobe function (e.g., Braver & Barch, 2002; West, 1996); (2) frontal damage is associated with impaired FOK accuracy (Janowsky, Shimamura, & Squire, 1989); and (3) older adults' FOK accuracy positively correlates with neuropsychological measures of frontal function (Perrotin et al., 2008; Souchay et al., 2000).

One explanation for the episodic FOK deficit in resolution is that older adults find it difficult to make accurate evidence-based evaluations of accessible information that may be relevant to later target recognition. We shall refer to this argument as the *inferential deficit hypothesis*. It parallels arguments that age deficits in inferential mechanisms may be responsible for age deficits in source monitoring (Henkel, Johnson, & DeLeonardis, 1998; Mitchell & Johnson, 2009), illusory overconfidence in recognition memory outcomes (Dodson, Bawa, & Krueger, 2007), and failure to control retrieval so as to generate diagnostic evidence about the desired target (e.g., Jacoby, Bisahara, Hessels, & Toth, 2005).

However, other evidence weighs against an inferential deficit explanation. MacLaverly and Hertzog (2009) found no age differences in episodic FOK resolution, and neither older nor younger adults' resolution was affected by delaying the FOK so that it was not made immediately after an explicit cued recall attempt.

Older adults are often reported to have a deficit in recollection during recognition tests (see, e.g., Mäntylä, 1993; Perfect & Dasgupta, 1997; see Light, Prull, La Voie, & Healy, 2000, for a review). Souchay et al. (2007) linked recollection deficits in older adults, as measured by remember-know (RK) judgments, to reduced FOK magnitudes. Age-related deficits in recollection are correlated with age-related changes in hippocampal activation (e.g., Daselaar, Fleck, Dobbins, Madden, & Cabeza, 2006) rather than with strategic or inferential deficits that are often associated with frontal functioning (e.g., Shing, Werkle-Bergner, Li, & Lindenberger, 2008). Recently, Daniels, Toth, and Hertzog (2009) found atypical age differences in the resolution of judgments of learning, in which resolution involved predicting reported recollection

experiences during a subsequent recognition test. Such findings suggest that age deficits in recollection (i.e., a memory deficit) may impair the resolution of metacognitive judgments, including episodic FOKs, for recognition tests under some conditions.

We propose that a valid test of the inferential deficit hypothesis for FOKs needs to address an alternative memory-related explanation of age-related episodic FOK deficits: the *memory constraint hypothesis*. According to this hypothesis, the age deficits in FOK resolution observed by Souchay et al. (2007) and in other studies could merely be an outcome of low levels of experimentally induced episodic memory strength in the older adult sample. Low memory strength could reduce the quality of information that is available and accessible when making the FOK, which in turn could limit FOK resolution even if aging leaves inferential processing largely intact. Before explaining the memory constraint hypothesis in further detail, we will offer a theoretical perspective on FOK resolution, grounding it in existing, relevant theories about how FOKs are formed.

### Conceptual Perspectives on the Construction of FOKs and FOK Resolution

The major current views regarding FOK construction treat FOK resolution as an outcome of heuristics based on people's accessing specific sources of information. All extant theories of FOKs reject the idea that individuals have direct access to information held in memory. Instead, FOKs can access only the products of cognition, including memory retrieval searches, although the feeling of knowing may involve an intuitive mechanism that arises from nonconscious influences deriving from target accessibility (Koriat, 2000; Metcalfe, 2000). According to Koriat (1993, 1995, 2000), FOKs are constructed judgments that are made on the basis of accessible information,<sup>2</sup> irrespective of whether that information is a valid predictor of subsequent recognition memory (see also Krinsky & Nelson, 1985; T. O. Nelson, Gerler, & Narens, 1984). A major competing hypothesis is that individuals make FOKs on the basis of the familiarity of the cue that is used to generate the FOK (Metcalfe, Schwartz, & Joaquim, 1993; Miner & Reder, 1994). For example, FOK magnitudes can be increased by subliminal priming of the FOK cue, even though this effect is irrelevant to whether information about the target is accessible in memory (see, e.g., Jameson, Narens, Goldfarb, & Nelson, 1990). In some situations, cue familiarity may be the primary source of influence on FOKs (Reder & Ritter, 1992). For instance, Koriat and Levy-Sadot (2001) argued that unfamiliar cues will generate a low FOK and a decision not to search memory; in this case, familiarity alone drives FOK judgments. By contrast, more familiar cues will cause people to search memory for information about targets, at which point the nature of the information accessed will drive FOKs. Evidence exists that both cue familiarity and information accessibility influence FOKs (e.g., Hosey, Peynircioğlu, & Rabinovitz, 2009).

Our perspective acknowledges and embraces both heuristics as sources of influence on FOKs, yet it also empha-

sizes the importance of accounting for variation in FOK resolution. Tests of the accessibility and cue familiarity accounts have typically evaluated FOK magnitudes and latencies rather than resolution (e.g., Koriat & Levy-Sadot, 2001; Metcalfe et al., 1993).

Consistent with the memory constraint hypothesis, FOK resolution increases when the quality of original encoding increases. Lupker, Harbluk, and Patrick (1991) demonstrated that an incidental sentence-generation orienting task produced much higher FOK resolution with recognition memory than did a shallow vowel-counting orienting task. They also demonstrated a benefit of intentional versus incidental learning instructions, with the provision of extra study time after intentional instructions having a differential benefit on resolution. Sacher, Taconnat, Souchay, and Isingrini (2009) recently showed that divided attention during encoding reduced episodic memory, FOK magnitudes, and FOK resolution. Nelson and colleagues (Carroll & Nelson, 1993; T. O. Nelson, Leonesio, Shimamura, Landwehr, & Narens, 1982) manipulated the degree of original overlearning of PA items, showing that FOK resolution increased as a function of increased levels of overlearning. Such outcomes implicate the quality of memory encoding as an influence on FOK resolution.

The basic principle we promote is that accessibility involves a mixture of diagnostic and nondiagnostic information about the underlying information available in memory, and that the manipulation of encoding quality increases the proportion of diagnostic information regarding the status of information in memory that can be accessed. Associative memory strength, even in the case of failed cued recall, derives from the quality of initial encoding of the new association, as well as the match between encoded features and demands of the recognition test (see, e.g., Jacoby et al., 2005). Resolution is increased when (1) the degree of diagnostic information accessed about the association is higher for some items than for others, and (2) people can base FOKs on these differences in information that derive from underlying associative memory strength. This conceptualization preserves the proposal that FOKs are heuristic in nature, but emphasizes that the diagnosticity of the information accessed is influenced by encoding quality. Resolution is increased because, on a probabilistic basis, accessing diagnostic information for some items—but not others—provides a valid basis for discriminating the probability of later recognition memory success.

We conceptualize accessibility of diagnostic information as stemming from a continuum of available evidence about the original encoding episode—evidence that can be reconstructed by self-initiated retrieval search or automatically triggered by presentation of a FOK cue. Items that are below a successful recall test threshold are graded in the amount of available and accessible information about the item and its encoding context stored in memory (Bower, 2000). From this point of view, the level of subthreshold memory strength of a given item, when cued during PA recall, activates information that

can be used to produce accurate FOKs. Encoding of PA items generates memory traces that include information about temporal context and activated associations of the words themselves (see, e.g., D. L. Nelson, Fisher, & Akirmak, 2007; Sederberg, Howard, & Kahana, 2008). This information is bound to the episode and could, in turn, influence FOKs if presentation of the FOK cue primed the retrieval of the associated information. A high FOK that precedes successful recognition may involve accessing distinctive, valid information about the target from accessible sources, despite the fact that the target cannot be retrieved (Koriat, Levy-Sadot, Edry, & de Marcas, 2003). Noncriterial recollection (e.g., Parks, 2007) involves the explicit recall of information about the encoding episode other than the target itself; it may be one of the principal bases of diagnosticity for FOKs. For example, if an individual used an interactive imagery mediator at study (see Richardson, 1998), he or she might successfully retrieve the interactive image but fail to reconstruct the correct target from it (Dunlosky, Hertzog, & Powell-Moman, 2005). In such cases, the explicit high-confidence recollection of the image would serve as a diagnostic cue from which the target could later be successfully recognized. Resolution will be enhanced when items vary in the accessibility of diagnostic information, and our argument is that the principal influence on this source of variation is the quality of encoding.

The notion of gradations in the quality of valid, accessible information for unrecalled items is supported by empirical demonstrations that high FOKs for unrecalled items are more likely to be followed by a recollective experience as measured by RK judgments for the subsequent recognition test (Hicks & Marsh, 2002; MacLaverly & Hertzog, 2009). One can argue that traditional FOK resolution for recognition accuracy shows a discrimination of low versus high subthreshold memory strength by FOKs, but that the FOK correlation with RK judgments shows graded access to valid sources of information about underlying memory strength for items above the recognition threshold (Bower, 2000). FOKs for successfully recognized items are just as likely to produce “remember” responses (MacLaverly & Hertzog, 2009) as when both successful and unsuccessful recognition trials are included, suggesting that FOKs are sensitive to information that forecasts gradations in recognition memory phenomenology. Note that above-chance FOK/RK correlations for successful recognition trials rely on gradations of diagnostic information that are, in effect, ignored when FOKs are validated solely against recognition success versus failure.

### **Contrasting the Memory Constraint and Inferential Deficit Hypotheses**

We are now ready to define the memory constraint hypothesis as an account of age differences in episodic FOK resolution. Given age-related deficits in episodic memory performance (see, e.g., Hoyer & Verhaeghen, 2006; Kausler, 1994; Zacks, Hasher, & Li, 2000), the typical single, experimenter-paced exposure to a list of PA items produces substantial age differences in memory perfor-

mance. Assuming that above-chance resolution of FOKs requires access to valid partial information about the cue–target association in memory when cued recall fails to recover the target, it could be the case that any age difference in FOK resolution reflects older adults' more impoverished memory representations of the original items based on limited initial encoding, which may arise from deficient encoding strategies (Dunlosky & Hertzog, 2001; Luo, Hendricks, & Craik, 2007), age deficits in associative binding mechanisms (Howard, Kahana, & Wingfield, 2006), or both (Naveh-Benjamin, Brav, & Levy, 2007). Older adults may have intact inferential mechanisms required for FOKs, but these mechanisms may be data limited (Norman & Bobrow, 1975) by their access to low-quality memory representations of the original episodes. Inferences would be starved for diagnostic information that could be used to generate a valid FOK. Consistent with the position of Dunlosky and Metcalfe (2009), we argue that tests of age differences in mechanisms for generating FOKs and FOK resolution should equate age groups on underlying memory performance, something that has rarely been done.

The memory constraint hypothesis predicts that if older and younger adults are equated on recall and recognition memory performance, FOK resolution will be equivalent for the two age groups. In contrast, the inferential deficit hypothesis predicts that age-related deficits in FOK resolution arise from impaired inferential processes that are critical for making accurate FOK judgments. Hence, the inferential deficit hypothesis predicts age deficits in episodic FOK resolution even when older adults are equated on memory task performance. To evaluate these competing hypotheses, we created a repetition delay task—inspired by T. O. Nelson et al.'s (1982) overlearning experiments—to equate the age groups on overall memory performance.

### The Repetition Delay Paradigm

In the repetition delay paradigm, individuals were shown PA items once, twice, or four times during an original study phase of the experiment. To maximize the probability of effective encoding, they were given descriptions of associative mnemonics such as interactive imagery and sentence generation (Bower, 1970; Richardson, 1998) prior to study. These mnemonics are highly effective for both young and old adults for learning new associations between normatively unrelated word pairs, although robust age differences in PA recall prevail even when they are used (see, e.g., Dunlosky & Hertzog, 1998; Price, Hertzog, & Dunlosky, 2008). After the encoding session, a second delayed session included PA (cued) recall, FOKs, and associative recognition memory tests. To better equate age groups on PA recall during this session, we imposed different delays after the encoding phase, giving longer delays to younger adults. Older adults were tested with either a 30-min or a 48-h delay; younger adults were tested with a 7-day delay. The 7-day delay was needed to avoid ceiling effects in recognition memory for repeated items, because a pilot experiment indicated that younger adults who were instructed to use

interactive imagery were at a recognition ceiling for repeated items with a 48-h delay.

To foreshadow, we successfully managed to create similar levels of memory performance for older adults with a 48-h delay and for the younger adults with the 7-day delay, making it possible to competitively evaluate the inferential deficit hypothesis and the memory constraint hypothesis. The inferential deficit hypothesis predicts that age-related deficits in FOK resolution will occur even when levels of memory performance are similar. The memory constraint hypothesis predicts instead that no age differences will arise in FOK resolution when memory performance is equated. Just as important, the memory constraint hypothesis also predicts that similar effects of repetitions on memory will produce similar increases in FOK resolution for both age groups.

Finally, another advantage of manipulating the number of item presentations is that, according to the graded memory strength view, repeated item exposure should manipulate gradations of memory strength both above and below the cued recall threshold, thereby influencing the level of memory performance, the level of FOK magnitude, and the resolution of FOKs. We hypothesized that repeated exposure to items during encoding would influence the correlation of FOKs for successfully recognized items with recognition CJs.<sup>3</sup> Taken together, repetition effects on FOK resolution for both recognition memory accuracy and recognition CJs would demonstrate that repetitions manipulate subsequent access to diagnostic information about underlying states of memory. If both young and old age groups showed similar benefits of repetitions on FOK resolution with recognition memory and CJs, it would provide even more compelling evidence for the memory-based view of FOK resolution and against a general inferential deficit in FOK accuracy for older adults.

## METHOD

### Participants

There were a total of 163 participants in this study. Younger adults participated in two sessions that were separated by a 7-day delay, whereas older adults participated in either a 30-min or a 48-h delay between session one and session two. The older adults in the 30-min delay condition were community-dwelling adults who were recruited from a participant pool of individuals residing in and around metropolitan Atlanta. Older adults in the 48-h delay condition were a combination of individuals from the metropolitan Atlanta pool and community-dwelling older adults from a volunteer pool near Kent State in Ohio. All of the older adults were compensated \$45 for participation. Younger adults, who were undergraduates at the Georgia Institute of Technology, could choose to receive either course credit or \$35 in monetary compensation. Table 1 reports relevant information about the participants, including the number of participants in each condition, mean age, cognitive test scores, and education level. We found typical patterns of age differences in the cognitive variables.

### Materials

**Experimental task.** This study was conducted on standard PCs with the experimental portions programmed using Visual Basic.Net (Microsoft Visual Studio, 2007). One list of 60 unrelated concrete noun–noun word pairs (e.g., DOG–SPOON) was created. All of the



**Table 1**  
**Participant Characteristics**

Variable	Age Group					
	Young 7 day ( <i>n</i> = 54; 46% Female)		Old 30 min ( <i>n</i> = 55; 73% Female)		Old 48 h ( <i>n</i> = 54; 61% Female)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Chronological age	19.56	1.19	68.32	6.09	69.76	6.26
Years of education	14.85	1.03	16.44	2.45	15.82	3.53
Advanced Vocabulary Test	17.83	4.55	18.45	8.22	21.78	7.72
Pattern comparison	23.50	3.41	14.49	3.19	14.81	3.20
Letter comparison	13.35	2.38	8.42	2.08	8.42	2.47
Raven's Matrices	19.00	3.32	5.65	3.83	6.15	3.85

words were selected from the University of South Florida Free Association Norms (D. L. Nelson, McEvoy, & Schreiber, 1998). The ListChecker Pro 1.2 program (Eakin, Schreiber, & Nelson, 2005)<sup>4</sup> was used to ensure that there was no semantic relatedness among the selected words. To construct the four-item forced choice associative recognition test, 120 additional never-studied targets were used as foils accompanying the original targets. These foils were also unrelated to the cues of the studied list. At study, the word pairs were presented one at a time, in a large, black sans serif font that was centered on a light gray background.

**Additional materials.** Participants completed a short demographic questionnaire in which they reported their age, education, and other variables, including a subjective rating of their current health status. Paper and pencil versions of the pattern comparison and letter comparison tasks (Salthouse & Babcock, 1991) were used as measures of perceptual speed. They also completed the Advanced Vocabulary Test (Ekstrom, French, Harman, & Dermen, 1976)—a recognition vocabulary test used to assess verbal ability. In addition, participants completed the Raven's Advanced Progressive Matrices measure (Raven, 1965), but with an imposed time limit. Finally, participants took a computer version of the Personal Beliefs about Memory Instrument (PBMI; Lineweaver & Hertzog, 1998). These data are not reported in the present article.

### Design and Procedure

In the first session program, participants studied the 60 word pairs one at a time, in the center of the screen, for 8 sec. Items were randomly assigned to three levels of repetition as a within-subjects factor; 20 items appeared once, twice, or four times during the study phase, under the constraint that items could not be repeated consecutively. Younger adults came back after 7 days for the memory tests (Session 2), whereas one group of older adults was tested the same day following a 30-min delay after completing the paper-and-pencil measures and the PBMI. The second group of older adults was tested 48 h later.

Before studying the PA items, participants were given a short practice list in order to become acclimated to the task environment. They were informed and given examples of three possible study strategies that they could employ to study the word pairs (rote repetition, sentence generation, interactive imagery). Participants saw a short practice, during which they made strategy reports after the presentation of each item.

Immediately after studying each PA item, participants made a strategy report using the following scale: 1, *used rote repetition*; 2, *used sentence generation*; 3, *used interactive imagery*; 4, *used rote repetition and either sentence generation or interactive imagery*; 5, *used other strategy, one not listed here*; 6, *tried to use a strategy but could not implement it*; 7, *did not attempt to use a strategy*. Participants had been instructed that they would see repeated presentations of some items. They were also told that they were free to change their study behavior if desired; that is, they were free to study an item with a different strategy than the one they utilized in an earlier presentation.

After the study phase, all participants completed the paper-and-pencil measures listed earlier. Older adults in the 30-min delay condition began Session 2 immediately following a short break; the other two groups returned for Session 2 after the specified delay.

Session 2 began with instructions about the memory tests and metacognitive judgments. During cued recall, participants saw the first word of a previously studied pair—for example, DOG-??—and were to type in the second word that was studied with the word that was shown. Following the participant's response, regardless of whether the answer was correctly recalled or not, the participant was instructed to make an FOK: "How confident are you from 0 to 100 that you will be able to recognize the word that was paired with the word shown above, if shown both words together?" They scaled their judgments on the scale from 0 = *definitely won't recognize* to 100 = *definitely will recognize*. Note that having participants make FOK judgments for all items has a precedent in the field (see, e.g., Koriat, 1993). Delaying FOKs by testing them in a block after all recall attempts was found by MacLaverly and Hertzog (2009) to produce equivalent effects on judgment magnitude and resolution as compared with when FOKs were given immediately after the recall attempt.

Participants were then given a four-alternative forced choice recognition task. They were presented with four sets of word pairs, requiring them to select the previously studied (intact) pair. The alternatives paired the original cue with (1) the correct target, (2) a previously studied target of another pair, or (3) two never-studied foils. Alternatives were randomly assigned to a position in the test list. After each recognition response, participants made a CJ: "How confident are you on a scale from 0 to 100 that the answer you just provided was correct?" where 0 = *not at all confident* and 100 = *completely confident*. After recognition testing concluded, the participants were thanked, debriefed, and compensated for their participation.

## RESULTS

All statistical analyses of means were conducted using a mixed-model analysis with SAS PROC MIXED, which uses restricted maximum-likelihood estimation of parameters for (possibly) incomplete repeated measures designs, assuming that any missing data are missing at random. This feature is especially useful when gamma correlations are the dependent variable, because extreme marginal distributions of FOKs or of recognition memory (such as near-ceiling levels of recognition performance) can render gammas noncomputable, creating missing values for certain cells. Traditional repeated measures analysis methods require complete data for all cases, so cases with incomplete data would be removed from the analysis. By retaining all available information, the

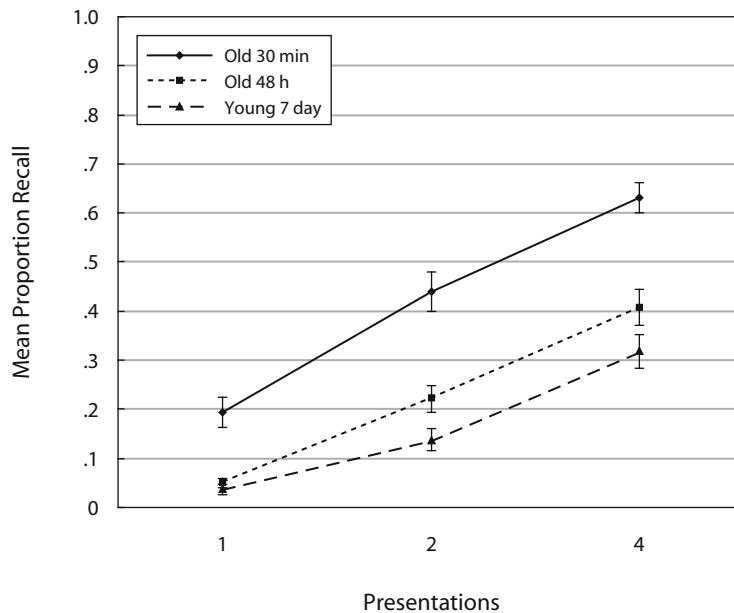


Figure 1. Cued recall as a function of presentations.

mixed-model analysis maximizes the power of the subsequent statistical tests (see Hoffman & Rovine, 2007, for further discussion of mixed-model applications for psychological experiments). All of the models were estimated assuming an unrestricted covariance matrix of the error terms across repeated measures (see Littell, Milliken, Stroup, Wolfinger, & Schabenberger, 2006). We report effect sizes using an extension of Cohen's (1988)  $d$  statistic that expresses least-squares fitted mean differences as a function of the appropriate pooled error term,  $d^* = (M_1 - M_2)/\text{SQRT}(\text{pooled variance estimate})$ . It can be interpreted as the number of standard deviations separating the means in question. Cohen suggested benchmarks of 0.2, 0.5, and 0.8 for small, medium, and large effect sizes, respectively.

### Recall and Recognition Performance

Figure 1 shows the cued recall data for all three groups as a function of the number of presentations. As expected, the repetition of items at encoding affected cued recall accuracy in all three groups [ $F(2,160) = 194.43, p < .001$ ]. The increase from one presentation to four presentations produced a large effect size ( $d^* = 1.69$ ). The three delay groups also differed in recall level [ $F(2,160) = 25.46, p < .001$ ]. The level of recall was slightly higher for the 48-h-delay older adults ( $M = .23, SE = .03$ ) than for the 7-day-delay younger adults ( $M = .16, SE = .03$ ) [ $F(1,106) = 4.08, p < .05$ ] ( $d^* = 0.32$ ), and cued recall was reliably higher for the 30-min-delay older adults than for the 7-day-delay younger adults ( $M = .42, SE = .03$ ) ( $d^* = 1.37$ ). The delay group interacted with the number of presentations, with larger linear increases in recall for the 30-min-delay older adults [ $F(4,160) = 5.76, p < .001$ ]. The recall difference in one versus four presentations just missed significance for the 48-h-delay older groups and

the 7-day-delay younger groups [ $F(2,106) = 2.95, p > .05$ ] (difference in  $d^* = 0.41$ ). In sum, recall performance was slightly better for the 48-h-delay older groups than for the 7-day-delay younger groups, but these small differences favoring older adults were actually better for testing the inferential deficit hypothesis.

Overall, recognition memory was higher than recall and was well equated by the different delays. Figure 2 shows the recognition data for unrecalled items only, which are traditionally the performance measures that are used to validate FOKs. The number of presentations had a major effect on recognition accuracy for unrecalled items [ $F(2,160) = 165.12, p < .001$ ]. The increase from one to four presentations generated a  $d^* = 1.51$ . For the recognition data, however, there was no main effect of delay group [ $F(2,160) = 2.24, p > .10$ ]. When restricting the comparison to 7-day-delay younger adults and 48-h-delay older adults, the marginal means were essentially identical (see Figure 2) [ $F(1,106) = 0.64, n.s.$ ] ( $d^* = 0.02$ ). The group  $\times$  repetition interaction [ $F(4,160) = 2.70, p < .05$ ] paralleled the cued recall data, showing greater repetition-based increases in the 30-min-delay older group and a negligible effect size when comparing only the 7-day-delay younger adults and 48-h-delay older adults in the repetition effect ( $d^* = 0.047, n.s.$ ).

In summary, recognition memory performance for the 48-h-delay older adults and for the 7-day-delay younger adults was successfully equated. Moreover, in standard aging designs, older adults almost always perform more poorly than younger adults in memory performance, so the present design created an atypical outcome of superior memory performance by the 30-min-delay older adults. An interesting question, relative to the memory constraint hypothesis, was whether these older adults would show FOK resolution superior to that of the other groups.

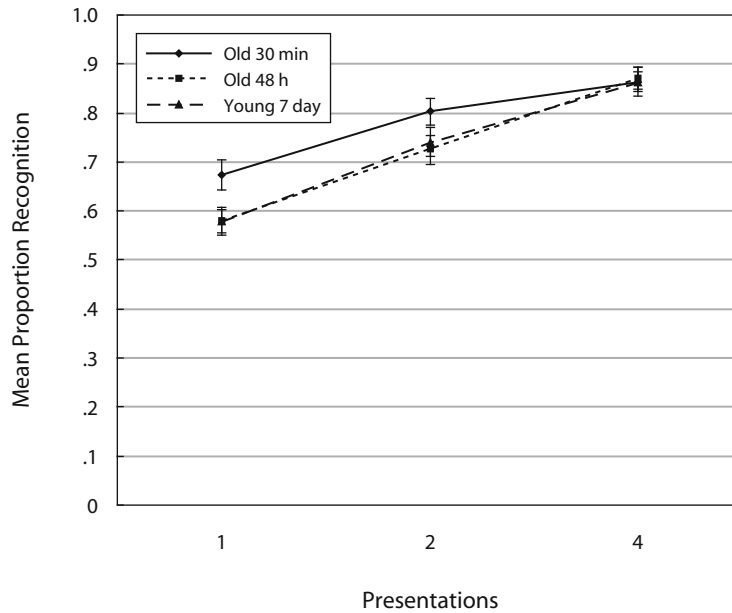


Figure 2. Four-alternative forced choice recognition: Unrecalled items only.

**Mean FOKs**

Although FOKs were collected for all items, maintaining an analog to the FOK experience requires evaluating them for unrecalled items only. On average, mean FOKs for unrecalled items increased reliably as the number of presentations increased (see Table 2) [ $F(2,160) = 72.42, p < .001$ ]. The main effect of repetition was also reliable when restricted to the critical 7-day-delay young group and to the 48-h-delay old group [ $F(2,106) = 50.75, p < .001$ ] ( $d^* = 0.56$ ). There were no reliable effects of group [ $F(2,160) = 0.47$ ] or a group  $\times$  delay interaction [ $F(4,160) = 1.34, p > .10$ , for both analyses]. In particular, the critical 7-day-delay young group and the 48-h-delay old group showed essentially equivalent mean FOKs ( $M_{\text{young}} = 28, M_{\text{old}} = 32$ ) [ $F(1,106) = 0.65, n.s.$ ] ( $d^* = 0.15$ ) and no group  $\times$  repetition interaction [ $F(2,106) = 1.58, p > .20$ ]. The difference in  $d^*$  for one to four pre-

sentations equalled 0.17. Mean FOK confidence for unrecalled items was low relative to the probability of successful recognition.

**FOK Resolution for Recognition Outcomes**

The critical data involved FOK resolution, measured by gamma correlations of FOKs for unrecalled items with later recognition memory performance. The resolution data were affected by some individuals' being at the recognition ceiling, especially with four presentations, limiting the number of persons with valid gamma correlations. This was less of an issue for the 48-hour-delay and 7-day-delay conditions (see Table 3).

Figure 3 shows that the number of presentations affected FOK resolution, with better resolution for items with repeated exposures [ $F(2,140) = 8.25, p < .001$ ]. On average, resolution for once-presented items was .08 ( $SE = .05$ ), a value not above chance [ $t(140) = 1.72, p > .05$ ]. With two and four presentations, resolution rose to .28 ( $SE = .06$ ) and .40 ( $SE = .07$ ), respectively, with both gammas reliably different from zero ( $p < .001$ ). The increase from one to four presentations produced a medium effect size ( $d^* = 0.48$ ). Hence, as expected, repeated opportunities to encode the items led to both better memory and better FOK resolution.

The group differences in resolution disconfirmed the inferential deficit hypothesis. With memory performance equated, no overall age differences occurred in FOK resolution between the 48-h-delay older adults ( $M = .19, SE = .06$ ) and the 7-day-delay younger adults ( $M = .20, SE = .05$ ) [ $F(1,99) = 0.01, n.s.$ ] ( $d^* = 0.01$ ). There was a reliable interaction of the two age groups with repetitions [ $F(2,99) = 3.91, p < .05$ ], with the older adult 48-h-delay group showing greater improvement in resolution with repeated exposure to the PA items (one to four presentation effect sizes

Table 2  
Mean Feeling of Knowing for Unrecalled Items

Presentations	Age Group					
	Young 7 day		Old 30 min		Old 48 h	
	M	SE	M	SE	M	SE
1	22.6	3.1	25.5	3.0	24.9	3.1
2	28.6	3.4	33.0	3.3	30.0	3.4
4	33.3	3.7	38.4	3.7	39.6	3.7

Table 3  
Valid Number of Cases in FOK-Recognition Gamma

Presentations	Age Group		
	Young 7 day	Old 30 min	Old 48 h
1	48	41	37
2	41	28	33
4	33	17	25

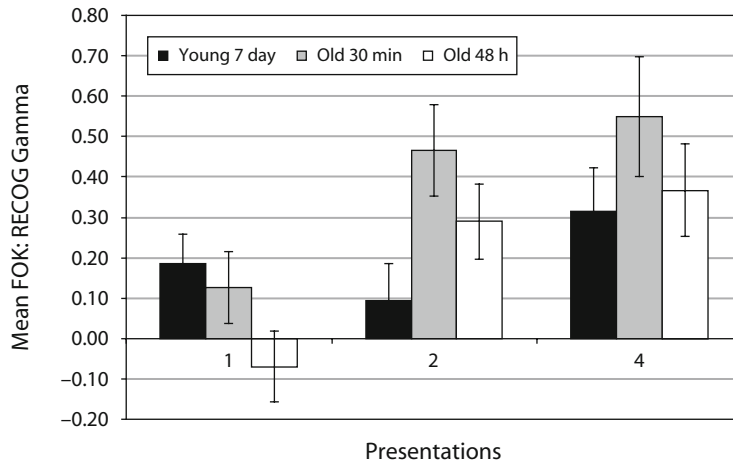


Figure 3. FOK: Recognition resolution.

of  $d^* = 0.22$  for young adults;  $d^* = 0.75$  for older adults). Young adults manifested better resolution for one-repetition items [unadjusted  $t(99) = 2.26, p < .05$ ] ( $d^* = 0.44$ ), but the greater gains in resolution for older adults reversed the direction of age differences in the two-repetition and four-repetition cell means. However, these age differences favoring older adults were not reliable ( $d^* = 0.32$  for two presentations;  $d^* = 0.09$  for four presentations).

Older adults in the 30-min delay condition showed, on average, superior resolution to (1) the young adult group [ $t(140) = 2.07, p < .05$ ] ( $d^* = 0.30$ ) and to (2) the older adult 48-h-delay group [ $t(140) = 2.13, p < .05$ ] ( $d^* = 0.32$ ). These effects are predicted by the memory constraint hypothesis, and, just as important, cannot be explained by the inferential deficit hypothesis. Older adults who were tested with a 30-min delay showed no deficit in resolution for single-presentation items, relative to the 7-day-delay young adults (the small effect,  $d^* = 0.10$ , favored the older adults), and showed a reliably larger effect of repetitions on resolution (difference in  $d^* = 0.47$ ).

### FOK Resolution With CJs

The gamma correlations of FOKs with CJs for all unrecalled items were greater than chance ( $p < .001$ ) and were influenced by the number of presentations [ $F(2,146) = 3.80, p < .05$ ]. The marginal mean gammas were .12 ( $SE = .03$ ), .23 ( $SE = .04$ ), and .27 ( $SE = .05$ ) for one, two, and four presentations, respectively. The effect size was small ( $d^* = 0.34$ ) between one-presentation and four-presentation items. Other effects were not reliable.

Figure 4 plots the gamma correlations for correctly recognized items only as a function of group and presentation condition. All of the values were above chance, indicating that differences in FOKs for items above the recognition threshold align with influences that generate confidence in the accuracy of forced choice recognition answers. The gamma correlations were reliably affected by the number of presentations [ $F(2,144) = 3.63, p < .05$ ], but not by delay group or the delay group  $\times$  presentation interaction ( $p > .25$ ). The marginal mean difference between the critical 7-day-delay young adults and the 48-h-delay old adults

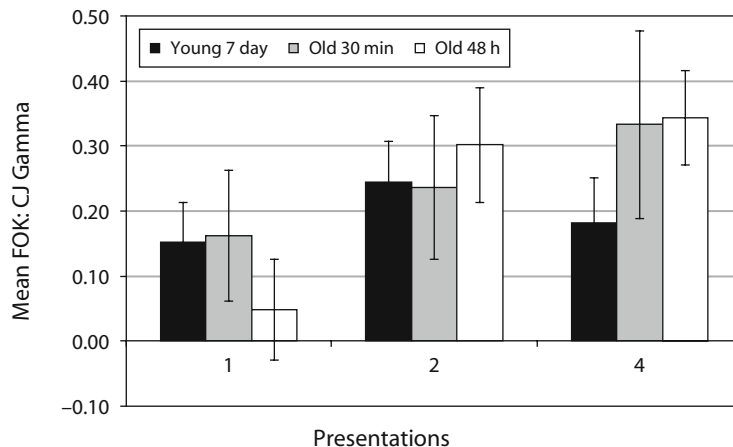


Figure 4. FOK: Retrospective confidence judgment resolution (recognized items only).



generated a negligible effect size ( $d^* = 0.08$ ). Despite similar trends in the sample means to the FOK recognition correlations (compare Figure 3), there was no meaningful difference between those two groups in FOK–CJ resolution for the once-presented PA items ( $d^* = 0.15$ ). The marginal mean gammas were .13, .25, and .29 (all  $SEs = .05$ ) for one, two, and four presentations, respectively. The difference between one and four presentations generated a medium effect size ( $d^* = 0.57$ ).

### Use of Effective Encoding Strategies and FOKs

Because we had collected item-level encoding strategy reports after individuals studied each PA item, we were able to evaluate whether the use of effective encoding strategies influenced FOKs. Our previous work has demonstrated strong relationships of self-reported use of effective mediators—interactive imagery, sentence generation, and other idiosyncratic mediator generation—on associative memory (Dunlosky & Hertzog, 2001; Hertzog, Dunlosky, & Robinson, 2009; Hertzog et al., 2010). As in these studies, we pooled reports of these three mediators into a normatively effective category, and the remaining options (rote repetition, none, ran out of time) were pooled into a normatively ineffective strategy category (for a rationale, see Bailey, Dunlosky, & Hertzog, 2009). With respect to repetitions, an item was assigned as having been encoded with an effective strategy if one was reported on any of the presentations.

**Rates of effective strategy production.** We first evaluated age differences and repetition effects on effective strategy use. On average, individuals reported using effective strategies (interactive imagery, sentence generation, or other; see Hertzog et al., 2010) for 80% of the items. This level of strategy use after being informed of their existence is consistent with other studies from our lab (e.g., Dunlosky & Hertzog, 2001). There was a trend for age differences in the use of effective strategies, with a main effect for delay group just missing significance [ $F(2,160) = 2.61, p < .08$ ]. Younger adults reported using effective strategies 86% of the time, and older adults reported using effective strategies 79% of the time ( $d^* = 0.25$ ); this slight age-related production deficiency is consistent with findings in the larger literature (see Dunlosky & Hertzog, 2001; Kausler, 1994). Repetitions influenced effective strategy use [ $F(2,160) = 54.11, p < .001$ ]. The use of effective mediators was least likely with a single presentation ( $M = 73.3\%$ ,  $SE = 2.2$ ), with mediator use increasing to  $M = 82.9\%$  ( $SE = 1.9$ ) with two presentations, and  $M = 87.6\%$  ( $SE = 1.8$ ) with four presentations. The difference between one and four presentations was a moderate-sized effect ( $d^* = 0.56$ ). There was no interaction of delay group with repetitions ( $F < 1$ ).

**Effective strategy use and FOKs.** If the quality of encoding influences subsequent FOKs, then self-reported use of effective strategies should influence FOKs. This would trivially be the case if we evaluated strategy reports' correlation with FOKs for all items, because of (1) the higher FOKs for correctly recalled items and (2) the often-demonstrated strong relationship of effective encoding strategies and the probability of cued recall.

Concerning the former effect, FOK confidence for correctly recalled items was considerably higher than FOK confidence for unrecalled items [ $F(1,160) = 788.87, p < .001$ ] ( $M_{\text{recalled}} = 82.5, M_{\text{unrecalled}} = 26.1$ ) ( $d^* = 2.52$ ). The principal question of interest here, therefore, was whether prior encoding history of an item influenced FOKs for items that could not be successfully retrieved during the cued recall test.

A mixed-model analysis of FOKs using repetitions, delay groups, and effective strategy use as predictors, revealed a small but reliable influence of effective strategy use on FOKs for unrecalled items [ $F(1,160) = 18.12, p < .001$ ] ( $d^* = 0.16$ ), independent of the reliable effect of repetitions. Mean FOKs for effective strategies ( $M = 31\%$ ,  $SE = 1.9$ ) were higher than mean FOKs for ineffective strategies ( $M = 27\%$ ,  $SE = 2.0$ ). There was also a trend for a repetition  $\times$  effective strategy use interaction [ $F(2,160) = 2.78, p < .10$ ]. FOK differences between effective and ineffective strategies were largest for the four-presentation condition ( $M_{\text{difference}} = 7\%$ ,  $d^* = 0.28$ ) versus the one-presentation condition ( $M_{\text{difference}} = 2\%$ ,  $d^* = 0.09$ ). In sum, this evidence supported the argument that quality of encoding, assessed at the item level, predicts FOK magnitudes for unrecalled items.

## DISCUSSION

Our discussion of these results is organized in two main sections. First, we review the findings with respect to the issue of the nature of underlying memory strength as an influence on FOK resolution. Second, we consider the issue of age differences in FOK resolution, contrasting the inferential deficit and memory constraint hypotheses in light of the findings of the present study.

### Effects of Repetitions of FOK Resolution

**FOK recognition resolution.** The present results show that manipulating the number of presentations of an item during study enhances the association of FOKs for unrecalled items with subsequent recognition memory accuracy, consistent with the results of earlier work on overlearning effects (Carroll & Nelson, 1993; T. O. Nelson et al., 1982) and other encoding quality manipulations (Lupker et al., 1991; Sacher et al., 2009). When the level of underlying memory strength was low (as in the one-presentation conditions in the present experiment), gamma correlations reflecting episodic FOK resolution were only marginally better than chance (zero correlation). As underlying associative memory strength increased due to repetitions of PA items at encoding, individuals were better able to forecast the likelihood of successful recognition. Furthermore, we demonstrated that the use of effective mediational strategies (such as interactive imagery) both enhanced memory and increased FOKs for unrecalled items. Thus, FOKs and FOK accuracy are influenced by the quality of the original processing of information at encoding.

These results are consistent with cue-familiarity and accessibility accounts of FOK construction (see, e.g., Koriat & Levy-Sadot, 2001), but they emphasize that the manipulation of underlying memory strength increases not

just the total amount of information that can be accessed, or the familiarity of the cue alone, but also either the type of information or the relative proportion of valid (diagnostic) information that can be accessed when engaging in a search for items that cannot be recalled.

An outstanding issue concerns whether underlying memory strength also influences the accuracy of semantic FOKs. Although speculative, we propose that the memory constraint hypothesis also holds for semantic FOKs. This proposal would seem consistent with the age equivalence found in semantic FOK accuracy, given that aging appears to leave semantic memory relatively spared (Balota, Dolan, & Duchek, 2000). Most importantly, this proposal can be empirically evaluated. In particular, if semantic knowledge was originally learned to differing levels of performance, differential semantic FOK accuracy should occur. For instance, Ackerman and Wolman (2007) showed that people demonstrate differentiated knowledge of domains that they originally studied in high school and college, and that they can, to a degree, accurately predict the relative likelihood of remembering information for different topics. This differential background knowledge could also affect FOK resolution for unrecalled items within a given domain. If they are in high school, it is possible that students learned facts about international politics much better than they did facts about agriculture. Therefore, FOKs made years later for unrecalled answers to these facts would be greater for international politics than for agriculture. Such effects would be logically, if not empirically, distinct from effects of topic familiarity on FOK magnitudes (e.g., Connor, Balota, & Neely, 1992).

**FOK–CJ resolution.** Item repetition also enhanced the correlation of FOKs with confidence (CJs) in the accuracy of correct recognition responses. This outcome extends previous findings from Hicks and Marsh (2002) and MacLaverly and Hertzog (2009) that showed reliable correlations of FOKs for RK judgments after a single encoding of each item. The FOK–CJ relationship suggests that influences on the degree of recollection in associative recognition, as assessed by CJs (see, e.g., Yonelinas, 2002), may also have graded influence on FOKs made prior to the recognition test. This argument is also supported by the outcome that repetitions at encoding increase these FOK–CJ correlations. We consider these reliable FOK–CJ correlations to be an important indication that gradations in memory strength influence the accessibility to diagnostic information about the underlying state of memory. Even though memory strength cannot be directly accessed when making an FOK (Koriat, 1995), there are apparently changes in the influence of accessible, diagnostic information on FOKs that vary as a function of original encoding conditions.

**Why does repetition increase FOK resolution?** The present study did not attempt to identify the specific sources of the diagnostic information accessed when FOKs were made, but we speculate that it is likely to be retrieval of information about the original cue–target encoding that does not enable target recovery itself. Retrieval of partial information about the desired target is an influence on FOK resolution (see, e.g., Koriat et al.,

2003); manipulating original encoding quality is likely to increase the probability of access to valid (diagnostic) partial information about the target.

In episodic FOKs, the partial information that is accessed may be a function of the type and quality of the encoding strategy that is used to form the new association. Dunlosky et al. (2005) instructed people to use sentence or imagery mediators to encode the type of unrelated PA items that were used in the present study. They showed that individuals can accurately report a gist-consistent description of their original imagery or sentence mediators but that they sometimes fail to decode the mediator so as to recover the correct target. Individuals also sometimes recollected partial information about a mediator, remembering the nature of the image or sentence but not all of its critical features. For example, if an individual had studied the item DOG–SPOON and had formed an interactive image of a dog on its haunches begging for food with a spoon in its mouth, he or she might remember the dog in a begging pose but not the spoon in its mouth. However, retrieval of this much information about the image might be diagnostic of successful recognition when confronted with SPOON as one of the forced choice alternatives. In general, the experience of noncriterial recollection (Parks, 2007) about the encoding episode, but not specifically about the target (e.g., whether one changed images upon repetition, whether one studied it early or late in the list), may be diagnostic of recognition, may increase with repetition and use of mediational strategies, and may be used as a basis for accurate FOKs. This hypothesis is testable, provided that future studies explicitly assess the mediator generated at study and then evaluate sources of information used at the time the FOK is generated (e.g., through think-aloud procedures).

### Aging and FOK Resolution

The demonstration that FOK resolution is affected by the nature and the quality of original encoding reinforces the concern that evaluations of the age-related inferential deficit hypothesis should be based on a comparison of age groups when memory performance is equated. Our experiment was successful in equating the 48-h-delay old adults with the 7-day-delay young adults in both recognition memory performance and mean FOKs for unrecalled items. Under these circumstances, we did not observe reliable age differences in FOK resolution for either recognition memory performance or for CJs for correctly recognized items when comparing the 48-h-delay old adults with the 7-day-delay young adults. Just as important, item repetition during study affected both groups' FOK resolution and had a slightly higher impact on older adults' FOK resolution. Finally, the 30-min-delay older adult group tended to have better FOK resolution than did either of the other two groups. These outcomes argue against a general inferential-deficit hypothesis. Instead, it appears that when underlying memory strength is sufficient to provide access to valid information about the target, older adults are both equally adept at forecasting successful recognition and equally likely to have FOKs discriminate variation in the confidence of correctly recognized targets.

The sole exception to this pattern of equivalent resolution for older adults was found with the single-presentation condition. There, younger adults' gamma correlations were reliably above chance after a 7-day delay (as was the case for older adults with a 30-min delay). In contrast, older adults with a 48-h delay—and with memory levels similar to those of the young adult group—did not manifest above-chance gamma correlations. The reliable age  $\times$  repetition interaction for these two groups showed that older adults' FOK resolution benefitted more than the younger adults' resolution when stimuli were repeated. The elimination of age differences in FOK resolution in the two-repetition and four-repetition conditions is not likely to be due to low statistical power to detect age differences, because the sample mean gammas for the older adults were actually higher than the corresponding gammas for the younger adult group.

The small but reliable age difference in FOK resolution in the single presentation condition is, in principle, consistent with a limited form of inferential deficit. Given that the inferential deficit and memory constraint hypotheses are not mutually exclusive, it is possible that when underlying memory strength is low and diagnostic information about targets is difficult to access, older adults are less likely to make accurate inferences about future item recognition. The single-presentation condition could be argued to be more representative of the type of encoding used in studies that have found age deficits in episodic FOK accuracy (e.g., Souchay et al., 2007), which may help reconcile the apparent differences in findings across studies. A limitation of the present study is that it did not fully map the surface of repetition-delay combinations for older and younger adults; future research doing so might help to identify the conditions under which inferential deficits might occur.

Consider the findings of Souchay et al. (2007) regarding the dissociation of age differences in resolution of semantic versus episodic FOKs. How can this effect be explained by the memory constraint hypothesis? We have already addressed the episodic resolution deficit for older adults. We can only speculate about the superiority of their semantic FOKs. Although there were no reliable age differences in the recall of the semantic information, older adults' superior semantic FOK resolution may be related to cohort differences in the degree of original exposure to the facts used by Souchay et al. (2007). That is, if older adults were more likely to have been repeatedly exposed to the information queried, then their FOKs for unrecalled information might have benefited to a greater degree from the recall of partial information about the target (e.g., Koriat et al., 2003). In this case, limited original encoding prior to the creation of context-free semantic knowledge would be more likely to constrain the resolution of the younger adults. In a related vein, Buchler and Reder (2007) argued that older adults often have more elaborated semantic networks that render them more susceptible to proactive interference effects when searching semantic memory. Interference would affect success rates of attempted recall of semantic information, but might still permit greater ac-

cess to partial information that would aid the resolution of older adults' semantic FOKs.

In any case, the elimination of any age-related inferential deficit in episodic FOKs by the within-subjects manipulation of repetition challenges the idea that age deficits in FOK resolution are associated with a stable individual-differences dimension in older adults, such as low frontal function (see, e.g., Perrotin et al., 2008). The same older adults who showed lower FOK resolution with a single presentation showed equivalent resolution when the stimuli were presented multiple times. Moreover, given that frontal patients have been reported to show deficiencies in strategic encoding behavior (e.g., Stuss & Alexander, 2005), findings of FOK resolution deficits in persons with low frontal function scores may be more consistent with their low-quality encoding than with inferential deficits at the time the FOK is made. Of course, such a conjecture would need to be evaluated in a study that directly measured frontal function and FOK resolution.

Another difference between the present study and its predecessors is that we informed all of the participants about the existence of strategies to encode the PA items. Older adults tend to have a mild production deficiency (Richardson, 1998), being less likely to spontaneously produce effective mediational strategies in PA tasks (see, e.g., Dunlosky & Hertzog, 2001; see Kausler, 1994). This tendency was manifested in the present study; the younger adults reported slightly higher rates of mediator use, even though the instructions promoted high levels of use in both age groups. Even larger differences exist when participants are uninformed about mediational strategies (e.g., Dunlosky & Hertzog, 2001; Hertzog et al., 2009). Given the effects of mediator use on memory and FOKs, it could be the case that age differences in episodic FOK resolution, when seen in the literature, are influenced by an age-related production deficiency in effective encoding strategies.

Although the present results do not fully rule out inferential deficits at low levels of underlying memory strength, they serve as a cautionary note regarding procedures that are appropriate for detecting such a deficit. Given the demonstration that the level of underlying memory strength influences the quality of FOK judgments, the best test of the inferential deficit hypothesis in FOK accuracy requires controlling age differences in the level of memory performance (Dunlosky & Metcalfe, 2009; Marquié & Huet, 2000). The conundrum is that some methods of equating memory performance, such as different item presentation times for persons of different ages, might introduce qualitative differences in underlying memory representations (e.g., by encouraging effective strategy use in older adults but discouraging it in younger adults; see Hertzog et al., 2009). Imposing different delays on young and old participants may be the best means of equating their memory performance, because forgetting is thought to be relatively passive and inert (but see MacDonald, Stigsdottir-Neely, Derwinger, & Bäckman, 2006) and does not involve a direct manipulation of original encoding behaviors that appear to influence FOK resolution. In the present study, when the



memory playing field was leveled by repeated stimulus presentation and different delays for old and young adults, older adults' episodic FOK accuracy was unimpaired, except at the lowest levels of memory performance.

#### AUTHOR NOTE

This research was supported by Grant R37-AG13148 from the National Institute on Aging, one of the National Institutes of Health (C.H., principal investigator). We thank Melissa Bishop, Teri Boutot, Bethany Geist, Ronit Greenberg, Devaki Kumarhia, Colin Malone, Melissa McDonald, Alisha Monteiro, Rory Murray, and Heather Bailey for assistance with data collection and data management. More information on research in the Hertzog laboratory can be obtained at <http://psychology.gatech.edu/CHertzog/>. Correspondence concerning this article should be addressed to C. Hertzog, School of Psychology, Georgia Institute of Technology, Atlanta, GA 30332-0170 (e-mail: christopher.hertzog@psych.gatech.edu).

#### REFERENCES

- ACKERMAN, P. L., & WOLMAN, S. (2007). Determinants and validity of self-estimates of abilities and self-concept measures. *Journal of Experimental Psychology: Applied*, *13*, 57-78. doi:10.1037/1076-898X.13.2.57
- ALLEN-BURGE, R., & STORANDT, M. (2000). Age equivalence in feeling-of-knowing experiences. *Journals of Gerontology*, *55B*, P214-P223.
- BÄCKMAN, L., & KARLSSON, T. (1985). The relation between level of general knowledge and feeling-of-knowing: An adult age study. *Scandinavian Journal of Psychology*, *26*, 249-258. doi:10.1111/j.1467-9450.1985.tb01162.x
- BAILEY, H., DUNLOSKY, J., & HERTZOG, C. (2009). Does differential strategy use account for age-related deficits in working memory performance? *Psychology & Aging*, *24*, 82-92.
- BALOTA, D. A., DOLAN, P. O., & DUCHEK, J. M. (2000). Memory changes in healthy young and older adults. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 395-410). New York: Oxford University Press.
- BENJAMIN, A. S., & DIAZ, M. (2008). Measurement of relative metamnemonic accuracy. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of memory and metamemory* (pp. 73-94). New York: Psychology Press.
- BOWER, G. H. (1970). Imagery as a relational organizer in associative learning. *Journal of Verbal Learning & Verbal Behavior*, *9*, 529-533. doi:10.1016/S0022-5371(70)80096-2
- BOWER, G. H. (2000). A brief history of memory research. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 3-32). New York: Oxford University Press.
- BRAVER, T. S., & BARCH, D. M. (2002). A theory of cognitive control, aging cognition, and neuromodulation. *Neuroscience & Biobehavioral Reviews*, *26*, 809-817. doi:10.1016/S0149-7634(02)00067-2
- BUCHLER, N. E. G., & REDER, L. M. (2007). Modeling age-related memory deficits: A two-parameter solution. *Psychology & Aging*, *22*, 104-121. doi:10.1037/0882-7974.22.1.104
- BUTTERFIELD, E. C., NELSON, T. O., & PECK, V. (1988). Developmental aspects of the feeling of knowing. *Developmental Psychology*, *24*, 654-663. doi:10.1037/0012-1649.24.5.654
- CARROLL, M., & NELSON, T. O. (1993). Effect of overlearning on the feeling of knowing is more detectable in within-subject than between-subject designs. *American Journal of Psychology*, *106*, 227-235. doi:10.2307/1423169
- COHEN, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- CONNOR, L. T., BALOTA, D. A., & NEELY, J. H. (1992). On the relation between feeling of knowing and lexical decision: Persistent subthreshold activation or topic familiarity? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *18*, 544-554. doi:10.1037/0278-7393.18.3.544
- DANIELS, K. A., TOTH, J. P., & HERTZOG, C. (2009). Aging and recollection in the accuracy of judgments of learning. *Psychology & Aging*, *24*, 494-500. doi:10.1037/a0015269
- DASELAAR, S. M., FLECK, M. S., DOBBINS, I. G., MADDEN, D. J., & CABEZA, R. (2006). Effects of healthy aging on hippocampal and rhinal memory functions: An event-related fMRI study. *Cerebral Cortex*, *16*, 1771-1782. doi:10.1093/cercor/bhj112
- DODSON, C. S., BAWA, S., & KRUEGER, L. E. (2007). Aging, metamemory, and high-confidence errors: A misrecollection account. *Psychology & Aging*, *22*, 122-133. doi:10.1037/0882-7974.22.1.122
- DUNLOSKY, J., & HERTZOG, C. (1998). Aging and deficits in associative memory: What is the role of strategy production? *Psychology & Aging*, *13*, 597-607. doi:10.1037/0882-7974.13.4.597
- DUNLOSKY, J., & HERTZOG, C. (2001). Measuring strategy production during associative learning: The relative utility of concurrent versus retrospective reports. *Memory & Cognition*, *29*, 247-253.
- DUNLOSKY, J., HERTZOG, C., & POWELL-MOMAN, A. (2005). The contribution of mediator-based deficiencies to age differences in associative learning. *Developmental Psychology*, *41*, 389-400. doi:10.1037/0012-1649.41.2.389
- DUNLOSKY, J., & METCALFE, J. (2009). *Metacognition*. Beverly Hills, CA: Sage.
- EAKIN, D. K., SCHREIBER, T. A., & NELSON, D. L. (2005). ListChecker Pro 1.2 [Computer software]. Available by contacting deakin@psychology.msstate.edu.
- EKSTROM, R. B., FRENCH, J. W., HARMAN, H. H., & DERMEN, D. (1976). *Kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.
- HART, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56*, 208-216. doi:10.1037/h0022263
- HENKEL, L. A., JOHNSON, M. K., & DELEONARDIS, D. M. (1998). Aging and source monitoring: Cognitive processes and neuropsychological correlates. *Journal of Experimental Psychology: General*, *127*, 251-268. doi:10.1037/0096-3445.127.3.251
- HERTZOG, C., DUNLOSKY, J., & ROBINSON, A. E. (2009). *Intellectual abilities and metacognitive beliefs influence spontaneous use of effective encoding strategies*. Unpublished manuscript.
- HERTZOG, C., KIDDER, D. P., POWELL-MOMAN, A., & DUNLOSKY, J. (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? *Psychology & Aging*, *17*, 209-225. doi:10.1037/0882-7974.17.2.209
- HERTZOG, C., SINCLAIR, S. M., & DUNLOSKY, J. (2010). Age differences in the monitoring of learning: Cross-sectional evidence of spared resolution across the adult life span. *Developmental Psychology*, *46*, 939-948.
- HICKS, J. L., & MARSH, R. L. (2002). On predicting the future states of awareness for recognition of unrecalable items. *Memory & Cognition*, *30*, 60-66.
- HINES, J. C., TOURN, D., & HERTZOG, C. (2009). Metacognitive influences on study time allocation in an associative recognition task: An analysis of adult age differences. *Psychology & Aging*, *24*, 462-475. doi:10.1037/a0014417
- HOFFMAN, L., & ROVINE, M. J. (2007). Multilevel models for experimental psychologists: Foundations and illustrative examples. *Behavior Research Methods*, *39*, 101-117.
- HOSEY, L. A., PEYNIROĞLU, Z. F., & RABINOVITZ, B. E. (2009). Feeling of knowing for names in response to faces. *Acta Psychologica*, *130*, 214-224. doi:10.1016/j.actpsy.2008.12.007
- HOWARD, M. C., KAHANA, M. J., & WINGFIELD, A. (2006). Aging and contextual binding: Modeling recency and lag recency effects with the temporal context model. *Psychonomic Bulletin & Review*, *13*, 439-445.
- HOYER, W. J., & VERHAEGHEN, P. (2006). Memory aging. In J. E. Birren & W. K. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 209-232). San Diego: Academic Press.
- JACOBY, L. L., BISHARA, A. J., HESSELS, S., & TOTH, J. P. (2005). Aging, subjective experience, and cognitive control: Dramatic false remembering by older adults. *Journal of Experimental Psychology: General*, *134*, 131-148. doi:10.1037/0096-3445.134.2.131
- JAMESON, K. A., NARENS, L., GOLDFARB, K., & NELSON, T. O. (1990). The influence of near-threshold priming on metamemory and recall. *Acta Psychologica*, *73*, 55-68. doi:10.1016/0001-6918(90)90058-N
- JANOWSKY, J. S., SHIMAMURA, A. P., & SQUIRE, L. R. (1989). Memory and metamemory: Comparisons between patients with frontal lobe lesions and amnesic patients. *Psychobiology*, *17*, 3-11.
- KAUSLER, D. H. (1994). *Learning and memory in normal aging*. San Diego: Academic Press.

- KORAIAT, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, **100**, 609-639. doi:10.1037/0033-295X.100.4.609
- KORAIAT, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General*, **124**, 311-333. doi:10.1037/0096-3445.124.3.311
- KORAIAT, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness & Cognition*, **9**, 149-171. doi:10.1006/ccog.2000.0433
- KORAIAT, A., & LEVY-SADOT, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 34-53. doi:10.1037/0278-7393.27.1.34
- KORAIAT, A., LEVY-SADOT, R., EDRY, E., & DE MARCAS, S. (2003). What do we know about what we cannot remember? Accessing the semantic attributes of words that cannot be recalled. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 1095-1105. doi:10.1037/0278-7393.29.6.1095
- KRINSKY, R., & NELSON, T. O. (1985). The feeling of knowing for different types of retrieval failure. *Acta Psychologica*, **58**, 141-158. doi:10.1016/0001-6918(85)90004-6
- LACHMAN, J. L., LACHMAN, R., & THRONESBURY, C. (1979). Metamemory through the adult life span. *Developmental Psychology*, **15**, 543-551.
- LIGHT, L. L., PRULL, M. W., LA VOIE, D. J., & HEALY, M. R. (2000). Dual-process models of memory in old age. In T. J. Perfect & E. A. Maylor (Eds.), *Models of cognitive aging* (pp. 238-300). New York: Oxford University Press.
- LINWEAVER, T. T., & HERTZOG, C. (1998). Adults' efficacy and control beliefs regarding memory and aging: Separating general from personal beliefs. *Aging, Neuropsychology, & Cognition*, **5**, 264-296. doi:10.1076/anec.5.4.264.771
- LITTELL, R. C., MILLIKEN, G. A., STROUP, W. W., WOLFINGER, R. D., & SCHABENBERGER, O. (2006). *SAS for mixed models* (2nd ed.). Cary, NC: SAS Institute.
- LUO, L., HENDRIKS, T., & CRAIK, F. I. M. (2007). Age differences in recollection: Three patterns of enhanced encoding. *Psychology & Aging*, **22**, 269-280. doi:10.1037/0882-7974.22.2.269
- LUPKER, S. J., HARBLUK, J. L., & PATRICK, A. S. (1991). Memory for things forgotten. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 897-907. doi:10.1037/0278-7393.17.5.897
- MACDONALD, S. W. S., STIGSDOTTIR-NEELY, A., DERWINGER, A., & BÄCKMAN, L. (2006). Rate of acquisition, adult age, and basic cognitive abilities predict forgetting: New views on a classic problem. *Journal of Experimental Psychology: General*, **135**, 368-390. doi:10.1037/0096-3445.135.3.368
- MACLAVERTY, S. N., & HERTZOG, C. (2009). Do age-related differences in episodic feeling of knowing accuracy depend on the timing of the judgment? *Memory*, **17**, 860-873. doi:10.1080/09658210903374537
- MÄNTYLÄ, T. (1993). Knowing but not remembering: Adult age differences in recollective experience. *Memory & Cognition*, **21**, 379-388.
- MARQUIÉ, J. C., & HUET, N. (2000). Age differences in feeling-of-knowing and confidence judgments as a function of knowledge domain. *Psychology & Aging*, **15**, 451-461. doi:10.1037/0882-7974.15.3.451
- MASSON, M. E. J., & ROTELLO, C. M. (2009). Sources of bias in the Goodman-Kruskal gamma coefficient measure of association: Implications for studies of metacognitive processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **35**, 509-527. doi:10.1037/a0014876
- METCALFE, J. (2000). Feelings and judgments of knowing: Is there a special noetic state? *Consciousness & Cognition*, **9**, 178-186.
- METCALFE, J., SCHWARTZ, B. L., & JOAQUIM, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 851-861.
- MICROSOFT VISUAL STUDIO (2007). Microsoft .Net Framework Version 3.5 SP1 [Computer software environment]. Redmond, WA: Microsoft.
- MINER, A. C., & REDER, L. M. (1994). A new look at feeling of knowing: Its metacognitive role in question answering. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 47-70). Cambridge, MA: MIT Press.
- MITCHELL, K. J., & JOHNSON, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin*, **135**, 638-677. doi:10.1037/a0015849
- NAVEH-BENJAMIN, M., BRAV, T., & LEVY, O. (2007). The associative memory deficit of older adults: The role of strategy utilization. *Psychology & Aging*, **22**, 202-208. doi:10.1037/0882-7974.22.1.202
- NELSON, D. L., FISHER, S. L., & AKIRMAK, U. (2007). How implicitly activated and explicitly acquired knowledge contribute to the effectiveness of retrieval cues. *Memory & Cognition*, **35**, 1892-1904.
- NELSON, D. L., MCEVOY, C. L., & SCHREIBER, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. Available at www.usf.edu/FreeAssociation/
- NELSON, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, **95**, 109-133. doi:10.1037/0033-2909.95.1.109
- NELSON, T. O., GERLER, D., & NARENS, L. (1984). Accuracy of feeling-of-knowing judgments for predicting perceptual identification and relearning. *Journal of Experimental Psychology: General*, **113**, 282-300. doi:10.1037/0096-3445.113.2.282
- NELSON, T. O., LEONESIO, R. J., SHIMAMURA, A. P., LANDWEHR, R. F., & NARENS, L. (1982). Overlearning and the feeling of knowing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **8**, 279-288. doi:10.1037/0278-7393.8.4.279
- NORMAN, D. A., & BOBROW, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, **7**, 44-64. doi:10.1016/0010-0285(75)90004-3
- PARKS, C. M. (2007). The role of noncriterial recollection in estimating recollection and familiarity. *Journal of Memory & Language*, **57**, 81-100. doi:10.1016/j.jml.2007.03.003
- PERFECT, T. J., & DASGUPTA, Z. Z. R. (1997). What underlies the deficit in reported recollective experience in old age? *Memory & Cognition*, **25**, 849-858.
- PERRONIN, A., TOURNELLE, L., & ISINGRINI, M. (2008). Executive functioning and memory as potential mediators of the episodic feeling-of-knowing accuracy. *Brain & Cognition*, **67**, 76-87. doi:10.1016/j.bandc.2007.11.006
- PRICE, J., HERTZOG, C., & DUNLOSKY, J. (2008). Age-related differences in strategy knowledge updating: Blocked testing produces greater improvements in metacognitive accuracy for younger than older adults. *Aging, Neuropsychology, & Cognition*, **15**, 601-626.
- RAVEN, J. (1965). *Advanced Progressive Matrices Sets I and II*. London: H.K. Lewis, 1965.
- REDER, L. M., & RITTER, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 435-451. doi:10.1037/0278-7393.18.3.435
- RICHARDSON, J. T. E. (1998). The availability and effectiveness of reported mediators in associative learning: A historical review and an experimental investigation. *Psychonomic Bulletin & Review*, **5**, 597-614.
- ROTELLO, C. M., MACMILLAN, N. A., REEDER, J. A., & WONG, M. (2006). The remember response: Subject to bias, degraded, and not a process-pure measure of recollection. *Psychonomic Bulletin & Review*, **12**, 865-873.
- SACHER, M., TACONNAT, L., SOUCHAY, C., & ISINGRINI, M. (2009). Divided attention at encoding: Effect on feeling-of-knowing. *Consciousness & Cognition*, **18**, 754-761. doi:10.1016/j.concog.2009.04.001
- SALTHOUSE, T. A., & BABCOCK, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, **27**, 763-776. doi:10.1037/0012-1649.27.5.763
- SCHACTER, L. (1983). Feeling of knowing in episodic memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 39-54. doi:10.1037/0278-7393.9.1.39
- SEDERBERG, P. B., HOWARD, M. W., & KAHANA, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review*, **115**, 893-912. doi:10.1037/a0013396
- SHING, Y. L., WERKLE-BERGNER, M., LI, S., & LINDENBERGER, U. (2008). Associative and strategic components of episodic memory: A life-span dissociation. *Journal of Experimental Psychology: General*, **137**, 495-513. doi:10.1037/0096-3445.137.3.495
- SOUCHAY, C., ISINGRINI, M., & ESPAGNET, L. (2000). Aging, episodic memory feeling-of-knowing, and frontal functioning. *Neuropsychology*, **14**, 299-309. doi:10.1037/0894-4105.14.2.299
- SOUCHAY, C., MOULIN, C. J. A., CLARYS, D., TACONNAT, L., & ISIN-



- GRINI, M. (2007). Diminished episodic memory awareness in older adults: Evidence from feeling-of-knowing and recollection. *Consciousness & Cognition*, **16**, 769-784. doi:10.1016/j.concog.2006.11.002
- STUSS, D. T., & ALEXANDER, M. P. (2005). Does damage to frontal lobes produce impairments in memory? *Current Directions in Psychological Science*, **14**, 84-88. doi:10.1111/j.0963-7214.2005.00340.x
- WEST, R. L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin*, **120**, 272-292. doi:10.1037/0033-2909.120.2.272
- YONELINAS, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory & Language*, **46**, 441-517. doi:10.1006/jmla.2002.2864
- ZACKS, R. T., HASHER, L., & LI, K. Z. H. (2000). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 527-539). Mahwah, NJ: Erlbaum.

#### NOTES

1. Gamma correlations have recently been criticized as a measure of resolution by Benjamin and Diaz (2008) and Masson and Rotello (2009), for different reasons. The latter authors preferred measures derived from signal detection theory, especially because of the benefits of computing indices of bias. Nevertheless, the gamma correlation is appropriate for

our purposes, and we use it because of its existence in other relevant research articles on FOK accuracy.

2. Koriat (1995) termed this *cue accessibility*, where cue refers to the information that is used to seed the FOK judgment. Given that FOKs are typically assessed with cued recall, where the cue initiates the search of memory for the target, we avoid semantic confusion by restricting use of the term *cue* to designate the recall cue, and we refer to evidence used to construct FOKs as *information*.

3. We used CJs to generate more continuously graded information about recognition phenomenology. Some studies suggest that RK judgments do not separate recollection from familiarity, but that they instead simply capture continuously graded information that may be better assessed by CJs (e.g., Rotello, Macmillan, Reeder, & Wong, 2006).

4. The ListChecker Pro 1.2 program uses the South Florida word association norms (Nelson et al., 1998) to make a variety of comparisons that are helpful in list construction, especially in regard to within-list and between-list associations. For more information on ListChecker Pro 1.2, contact Deborah Eakin, Assistant Professor, Department of Psychology, Mississippi State University, P.O. Box 6161, Mississippi State, MS 39762, 662-325-7949.

(Manuscript received July 20, 2009;  
revision accepted for publication February 11, 2010.)