

The relation between feelings of knowing and the number of neighboring concepts linked to the test cue

THOMAS A. SCHREIBER

University of Kansas, Lawrence, Kansas

and

DOUGLAS L. NELSON

University of South Florida, Tampa, Florida

We investigated whether feeling-of-knowing judgments are influenced by the number of different neighboring concepts linked to the test cue in long-term memory as measured using association norms. The purpose was to evaluate contrasting predictions made by the partial-retrieval hypothesis and the competition hypothesis. The partial-retrieval hypothesis assumes the more neighboring concepts activated by the test cue, the higher the feeling of knowing. In contrast, the competition hypothesis assumes that feelings of knowing are sensitive to competition between neighboring concepts, and it predicts that the fewer neighboring concepts activated by the cue, the higher the feeling of knowing. The findings were compatible with the competition hypothesis showing that both feeling-of-knowing and prediction-of-knowing ratings always were higher, the fewer different concepts were linked to the test cue. We obtained an identical pattern of results using different kinds of cues including taxonomic category names, ending sounds, and meaningfully related associates. We consider different ways that these findings could be reconciled with the partial-retrieval hypothesis, and we also discuss implications for other explanations of feeling-of-knowing effects.

How do people determine that they know information learned earlier even when they cannot recall that information at the moment? This question has been addressed by investigations of feeling-of-knowing (FOK) judgments showing that, when people cannot recall previously learned information, they nevertheless have a moderately accurate sense of whether they will be able to remember that information later (Hart, 1967; Leonasio & T. O. Nelson, 1990; T. O. Nelson, Gerler, & Narens, 1984; Schacter, 1983). The observation that people have some ability to predict whether they will be able to remember items later that they cannot remember now poses a theoretical challenge. If people do not base their FOKs on explicit access to the targeted information, then what is the basis for these feelings that enables them to predict future memory performance accurately?

One explanation is offered by the *partial-retrieval hypothesis* (e.g., Blake, 1973; Eysenck, 1979; Koriat, 1993; Koriat & Lieblich, 1977). Rather than basing FOK judgments on explicit access to the targeted information, this

hypothesis assumes that FOKs are based on related information that comes to mind during the course of searching memory. Such information can include candidate answers or "neighboring targets" that come to mind during the course of search, as well as lexical and semantic fragments of the target that come to mind. For example, when people have a tip-of-the-tongue (TOT) experience, they often have a very high FOK, and they can sometimes report linguistic attributes of the target word, such as the beginning letter or the number of syllables. Sometimes they also can report semantic attributes of the target as well as words that are related in meaning or sound (for a review, see Brown, 1991). On the basis of such observations, the partial-retrieval hypothesis assumes that FOKs are based on linguistic and semantic information coming to mind during the course of search, and the more such information coming to mind, the higher the FOK. This information is thought to be a source of positive evidence that at least some information about the target is known, and, presumably, it gives rise to a feeling that the target could be retrieved later even though it may not be retrievable now.

The main purpose of this investigation was to evaluate the assumption of the partial-retrieval hypothesis that FOKs should be higher, the greater the number of concepts that are linked to a test cue in memory. We used a relatively simple memory paradigm in which subjects were given extralist retrieval cues that had preexisting taxonomic, phonological, or associative relationships with the

This research was supported by NIMH Grant MH 16360 to the second author. The authors thank Tom Nelson, Asher Koriat, and Janet Metcalfe for their helpful comments on earlier versions of this article. Correspondence should be addressed to T. A. Schreiber, Department of Psychology, University of Kansas, 426 Fraser Hall, Lawrence, KS 66045 (e-mail: schreibe@kuhub.cc.ukans.edu).

—Accepted by previous associate editor Thomas O. Nelson

studied words but that were not presented during the study trial. One important characteristic of this *associative cuing paradigm* is that it is well suited for investigating the role played by preexisting memories in episodic memory tasks. Because the test cues were not presented during the study phase, subjects presumably must support their performance in part by relying on preexisting memories activated during the task (D. L. Nelson, Schreiber, & McEvoy, 1992). The key variable included in each experiment was *cue set size*, which refers to the number of preexisting connections that a stimulus, such as a word, has to other closely related concepts in long-term memory. For example, words with a relatively small number of preexisting connections to different meaningfully related words define relatively small associative sets, and words with more connections define larger associative sets. We measure set size using association norms (D. L. Nelson, Schreiber, & McEvoy, 1992; D. L. Nelson & Schreiber, 1992) and, when words with different numbers of preexisting connections to other words are used as test cues, cue set size is varied. Manipulations of cue set size based on normative data have been used to investigate the role played by preexisting memories in cued recall and other tasks, and we felt such manipulations could be used to explore assumptions of the partial-retrieval hypothesis.

Many experiments have shown that words connected to smaller associative sets are more effective cues for recall than are words connected to larger sets (for reviews, see D. L. Nelson, 1989, and D. L. Nelson, Schreiber, & McEvoy, 1992). Such results are obtained in the associative cuing paradigm in which subjects study individually presented words and are then given extralist retrieval cues during a recall test. In addition, target words can be recalled significantly more quickly given cues connected to smaller sets of associates relative to those connected to larger sets of associates (D. L. Nelson, McEvoy, & Bajo, 1988; Schreiber, 1993; Schreiber & Sergeant, 1998). Cue set size is associated with performance in recall using a variety of different kinds of cues (e.g., taxonomic category names, associates, ending and beginning stems, rhymes, word fragments, and pictures) and for different types of subjects (e.g., college students, elderly individuals, and children). These findings suggest that the inverse relation between set size and recall is relatively robust and is not limited to the use of cues from particular domains or subject populations.

A model called PIER¹ has been used to explain why set size is inversely related to recall (D. L. Nelson, Schreiber, & McEvoy, 1992). This model assumes that, when a stimulus such as a word is presented as a prompt to recall a related target word, preexisting connections to closely related concepts are activated or primed very rapidly and in parallel. The purpose of activating such prior knowledge is to provide rapid access to closely related concepts that could aid comprehension and that could potentially be relevant to performance in subsequent tasks (see Kintsch, 1988). In PIER, the activation of pre-

existing connections to related concepts is most likely to influence performance in a task when the test cue provides only partial information about the studied item, such as its membership in a taxonomic or rhyme category or its relation to an associated word. Under such conditions, PIER assumes neighboring concepts that are activated by the test cue can compete with the target. When there are fewer activated concepts in the pool, there is less competition with the target, and the probability of sampling the target is higher and the time to retrieve it is faster than when there are more activated concepts in the pool (see Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1983).

The specific purpose of the present experiments was to determine whether FOK judgments, like recall, would also be affected by the size of the associative set defined by the test cue. According to the partial-retrieval hypothesis, FOK ratings should be directly related to set size. Higher ratings should be given to cues that activate more related concepts in long-term memory than to cues that activate fewer related concepts. FOK ratings should be directly related to cue set size because more neighboring information should provide a greater sense that the target is known. A contrasting hypothesis is that FOKs, like recall, are sensitive to the amount of competition produced when related items are activated in long-term memory. The more competing items activated, the greater the amount of interference produced. As such, the competition hypothesis predicts that FOKs should be inversely, rather than directly, related to cue set size.

To evaluate these contrasting predictions, each experiment used the associative cuing procedure and varied cue set size as measured using association norms. The cues were of different types in different experiments, but the main manipulation always consisted of cue set size, with half of the test cues defining small and the remaining half defining larger sets of related items. In Experiment 1, taxonomic category cues were used to prompt recall. Immediately following the recall test, the ineffective cues were reshowed, and the subjects were asked to make FOK judgments for these cues. In Experiment 2, recall was prompted by cues varying in rhyme set size, and both prediction-of-knowing (POK) and FOK ratings were obtained from different groups of subjects. Experiment 3 replicated portions of Experiment 2 but used associatively related words as cues instead of category names to prompt recall. The purpose was to determine whether similar findings would be obtained for cues in which set membership is defined less clearly and for which category size information is less likely to be stored directly. Discussion of additional manipulations and rationale will be postponed until each experiment is introduced.

EXPERIMENT 1

The purpose of Experiment 1 was to demonstrate that there is a relationship between cue set size and FOKs. During study, the subjects were shown a list of individually presented words (e.g., SPOON or COLLIE). Following

study, they were given the names of taxonomic categories (e.g., EATING UTENSIL, DOG BREED), with instructions to use them as cues to recall the corresponding study words. Following standard procedures, during the FOK phase, the category names for words that the subjects could not recall were reshown, with instructions to make FOK ratings. The most important variable was taxonomic category size. Half of the taxonomic category cues defined relatively small sets of different instances, and the remaining half defined larger sets of instances. Category size was crossed with the number of study trials that the subjects received prior to test, in order to determine whether similar findings would be obtained under somewhat different conditions of target encoding.

Method

Design and Subjects. The design formed a 2×2 mixed-model factorial, with cue set size (small, large) manipulated within subjects and with study trials (one, two) manipulated between subjects. Thirty-two undergraduate students from the University of South Florida participated for course credit, and an equal number was randomly assigned to each study trial condition.

Materials. Appendix A shows the category name cues and associated targets for each level of cue set size. They were selected from taxonomic category norms (McEvoy & D. L. Nelson, 1982). In these norms, 169 participants were given the names of taxonomic categories and were asked to write down the first word that came to mind that was an instance of the presented category. These norms were used to measure functional category size by counting the number of different but appropriate instances given. A single-response, rather than multiple-response, technique was used to collect the normative data because it provides a better estimate of the functional set of items associated with a category and because it avoids problems associated with response chaining and retrieval inhibition (cf. Joelson & Herrmann, 1978; D. L. Nelson & Schreiber, 1992; D. L. Nelson, Schreiber, & McEvoy, 1992). The strength of the relationship between a category cue and instance was determined by dividing the frequency with which a particular instance was given by the number of subjects in the normative sample. Forty-two category names were selected from the norms, with an equal number defining relatively small and large sets of instances. These names served as extralist cues in the recall and FOK phases. The mean numbers of instances per category as measured using the norming procedure were, respectively, 6.24 ($SD = 1.51$) and 22.24 ($SD = 4.9$). Category-to-instance strength was equated for the two levels of set size, and the means were, respectively, .13 ($SD = .09$) and .12 ($SD = .02$). Hence, in the absence of study, the category cues produced the target approximately 12% of the time. Two instances from each category were selected to construct two separate lists of targets, and an equal number of subjects in each condition was randomly assigned to each list.

Procedure. The experiment involved study, cued recall, and FOK ratings. In the study phase, target words typed in uppercase letters were presented at a 3-sec rate using a Kodak carousel projector in an order that was unsystematically randomized for each subject. For the group receiving two study trials prior to test, the targets were presented a second time at the same rate and in the same order. All subjects were instructed to read each target word aloud and to remember as many words as possible. They were not told about subsequent phases of the experiment; however, they were told that some questions about the target words would be asked later, so they should concentrate hard on each word. To acquaint the subjects with the presentation rate, they were shown six first names prior to the study phase which they were asked to read aloud.

Following study, the subjects were given a cued recall test in which category names were presented one at a time. The subjects were instructed to read the name aloud and then to recall the related study word. The experimenter gave clarifying examples beforehand. The test was self-paced, and guessing was permitted when unsure. Following cued recall, the category names associated with unrecalled items were re-presented, and the subjects verbally indicated their FOK for the correct answer. Ratings were based on a 6-point scale, with 6 referring to the most positive feeling of knowing (100% certainty) and 1 referring to the least positive (100% uncertainty). In addition, the subjects were instructed to partition their overall FOK into six equal intervals. That is, they were asked to use the scale such that the difference between any adjacent points on the scale would be equal. They also were encouraged to use the entire range of numbers in computing their FOK such that it accurately reflected their FOK for the studied word. As with the recall test, the category names were presented individually; after reading each name aloud, the subjects repeated it aloud before responding. The instructions emphasized that only category names for unrecalled items would be shown. These unrecalled items included both omission and commission responses. Although the statistical analyses we will report included ratings for both kinds of responses pooled, we note that none of the patterns of findings were different when separate analyses were conducted for each type of recall response (e.g., see Table 2). Finally, items in each phase of the experiment were presented in a different random order for each subject.

Results and Discussion

The criterion for significance was set at .05 for all effects in all experiments. We will describe the findings for FOK ratings first, followed by the recall findings. Because none of the patterns of findings in any of the experiments in this series differed significantly as a function of the particular list of materials used or the type of recall error made, analyses were pooled across type of list and recall error to conserve space. In addition, for knowing ratings, we first computed each subject's median knowing rating for each condition, and these medians were then entered into a conventional analysis of variance (ANOVA). For the summary measure, we report the mean of each subject's median knowing ratings; however, note that the pattern of findings was identical when mean FOK ratings, instead of median ratings, were computed for each subject.²

Feelings of knowing. Figure 1 shows the mean of the subjects' median knowing ratings for unrecalled items as a function of cue set size and number of study trials. As indicated in this figure, FOK ratings were higher for category cues defining smaller sets of instances (4.09) than for those defining larger sets of instances (3.58). The statistical analysis indicated that this source of variance was significant [$F(1,30) = 11.32$, $MS_e = 0.38$]. This result suggests that the magnitude of FOK ratings is related to the number of category instances associated with a test cue, and that such ratings are higher when there are fewer instances associated with the cue than when there are many such instances.

Overall, ratings also were higher after two study trials (4.09) than after one (3.58), but the statistical analysis indicated that this difference did not quite reach the criterion for significance [$F(1,30) = 4.30$, $p < .06$]. This result

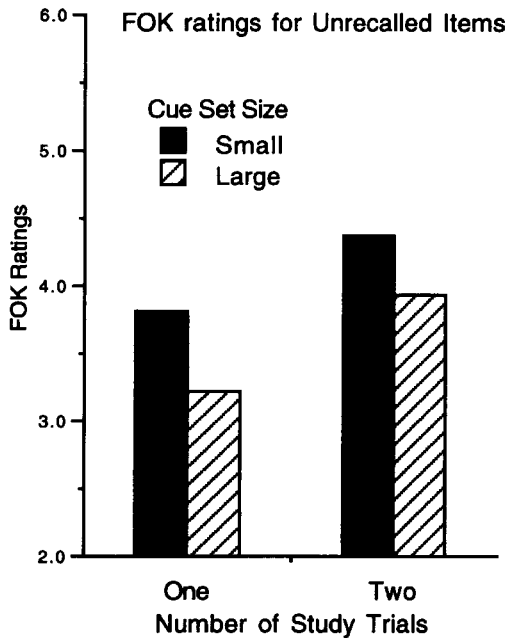


Figure 1. The mean of the individual subjects' median FOK ratings as a function of cue set size and number of study trials in Experiment 1. With taxonomic category names, cue set size is inversely related to the magnitude of FOK ratings.

is in general agreement with other findings suggesting that obtaining statistically significant effects related to degree of target encoding is difficult in the FOK task (e.g., Reder & Ritter, 1992; Schwartz & Metcalfe, 1992; for an exception, see Schreiber, 1998, in which different encoding strategies were compared). The interaction between set size and study trials was not significant with ($F < 1$).

Recall. The first row of Table 1 shows the mean probabilities of recall as a function of cue set size and number of study trials. Probability of recall was higher for cues defining smaller sets (.71) than for those defining larger sets (.56) [$F(1,30) = 28.13, MS_e = 0.01$]. Recall also was more likely following two study (.70) trials than following one (.57) [$F(1,30) = 6.93, MS_e = 0.04$], and there was no interaction between these sources ($F < 1$).

Each of these findings replicates a very large number of other studies showing similar effects of study trials and cue set size in recall (reviewed in D. L. Nelson, Schreiber, & McEvoy, 1992).

The most important result was that there was a relationship between the magnitude of FOK ratings and category size. On average, median ratings were higher for cues linked to smaller sets of instances than for those linked to larger sets of instances. These findings, based on the use of taxonomic category cues, are compatible with the inference that FOK ratings can be influenced by the amount of competition between the target and neighboring concepts that also are linked to the test cue. The findings appear to be inconsistent with at least one assumption made by the partial-retrieval hypothesis because it assumes that FOK ratings should be lower when fewer concepts are linked to the test cue, which is the reverse of the pattern that was obtained.

EXPERIMENT 2

The primary purpose of Experiment 2 was to extend the findings of Experiment 1 by employing cues linked to differing numbers of rhyming instances rather than taxonomically related ones. The main question was whether similar findings could be obtained for cues that provide access to information from a different domain. The cues were the ending sounds of words (e.g., /INE/ for the target word PINE). Half of the ending sounds defined relatively small rhyme categories, and the remaining half defined larger rhyme categories. During the testing phase, the subjects were shown the ending stems of words, and attention to the sound of the stem was emphasized. To focus the subjects' attention on sound, the experimenter pronounced the ending sound aloud on each trial, and the subjects were asked to repeat it aloud both before making their knowing rating and before providing a recall response.

Rhyme cues can be very effective cues in recall (e.g., D. L. Nelson & McEvoy, 1979; D. L. Nelson, Schreiber, & Holley, 1992), and the principle question was whether or not the number of rhymes related to a sound influences FOK ratings. In addition to varying rhyme set size, two types of FOK tasks were employed. The first task

Table 1
Mean Probabilities of Correct Recall for Experiments 1-3

Condition	One Study Trial		Two Study Trials	
	Small Cue Set	Large Cue Set	Small Cue Set	Large Cue Set
Experiment 1				
FOK	.65	.49	.77	.63
Experiment 2				
FOK	.58	.42	.72	.49
POK	.60	.35	.74	.49
Experiment 3				
FOK	.71	.59		
POK	.65	.49		

Note—FOK, feeling of knowing; POK, prediction of knowing.

was based on the standard FOK procedure in which subjects study targets, perform a cued recall test, and make FOK ratings on unrecalled items. This is the same task that was used in Experiment 1. In the second task, the order of the recall and rating phases was reversed, and, as a consequence, ratings were made on all items instead of only unrecalled items. This procedure permits analyses based on recalled and unrecalled items and thus provides additional information that cannot be obtained with the standard task. We wanted to determine whether a similar relation between knowing ratings and set size would occur for each kind of item. For ease of communication, the standard rating task will be referred to henceforth as the FOK task, and the second rating task in which ratings are made on all items will be referred to as the POK task. Versions of both tasks have been employed in prior research (e.g., T. O. Nelson & Narens, 1990; Reder & Ritter, 1992; Schreiber, 1993, 1998). Following prior use of the POK task, the subjects were encouraged to make their ratings as rapidly as possible, relying on their initial impressions. In contrast, FOK ratings were self-paced.

Method

Design and Subjects. The design formed a $2 \times 2 \times 2$ mixed-model design, with task (FOK, POK) and number of study trials (one, two) manipulated between subjects and with cue set size (small, large) manipulated within subjects. Sixty-four undergraduate students were drawn from the same sources, assigned to conditions, and rewarded for their participation in the same manner as in Experiment 1.

Materials. The ending stem cues and target words were selected from a normative database that was compiled by auditorially presenting the ending stems of words to large groups of participants ($n = 100\text{--}150$), who were asked to write the first word to come to mind that rhymed with each stem (see D. L. Nelson & McEvoy, 1979). The number of different, but appropriate, responses given in the normative sample was used to measure the size of the functional rhyme category defined by each ending sound, and the probability with which a particular response was given in the sample again was used to estimate the strength of the preexisting relationship between a sound and a particular response (i.e., cue-to-target strength). For example, using this procedure, the ending sound /URST/ defined a relatively small functional rhyme category of five instances, and the target word THIRST was given with a probability of .02. The pool of items comprising the normative database was used to prepare two separate lists of stems and targets. Each list was assigned to an equal number of subjects and is shown in Appendix B. There were 56 targets in each list, representing an equal number of items for each level of cue set size. When pooled across lists, mean cue set size for stems defining smaller rhyme categories was 6.47 instances ($SD = 1.57$); for stems defining larger rhyme categories, the mean was 20.62 instances ($SD = 5.11$). Mean cue-to-target strengths for the levels of cue set size were, respectively, .05 ($SD = .03$) and .04 ($SD = .03$). Relatively weak cues were used because rhyme cues tend to be very effective even when cue-to-target strength is relatively low (D. L. Nelson & McEvoy, 1979). Weaker cues were expected to produce intermediate levels of recall, which are desirable for producing relatively large pools of recalled and unrecalled items. A separate normative database was used to measure and equate associative set size of the target words for each level of cue set size because target set size influences recall even when

rhyme cues are used as test cues (e.g., D. L. Nelson, Schreiber, & Holley, 1992; Schreiber, 1998). The procedure for measuring associative set size is similar to the procedure for measuring the size of rhyme categories and is described in more detail in the Method section of Experiment 3. For the targets used in Experiment 2, associative set size of the target was equated for each level of cue set size, and the means were, respectively, 13.53 associates ($SD = 5.99$) and 13.62 associates ($SD = 5.56$).

Procedure. As with Experiment 1, this experiment involved study, cued recall, and knowing ratings. The study phase and instructions given to the subjects were the same as those used in Experiment 1. Again, target words were presented at a 3-sec rate using a Kodak carousel projector in a different random order for each subject; for the group receiving two study trials prior to test, the targets were presented a second time at the same rate and in the same order. The study instructions neither encouraged nor discouraged attention to word sound.

The procedure used in the FOK condition also was similar to that in Experiment 1 with the main exceptions relating to the use of rhyme cues instead of taxonomic category names. The subjects performed a cued recall test immediately following study and then made FOK ratings on unrecalled items. In recall, the ending stems were shown one at a time; to ensure attention to sound, the experimenter pronounced each stem aloud, and the subjects repeated it aloud on every trial before responding. The instructions indicated that, after pronouncing the stem aloud, the subjects should try very hard to recall the study word that rhymed with the stem. The experimenter provided an example of the procedure and the type of cue that would be given. Following cued recall, the stems for unrecalled items were reshown and, after hearing it and pronouncing it themselves, the subjects verbally indicated their FOK. Ratings were based on a 6-point scale, as with Experiment 1. The instructions emphasized that only stems for unrecalled items would be presented for FOK ratings.

The procedure used in the POK condition was similar to the procedure used in the FOK condition except that the order of the recall and rating phases was reversed, and ratings were made on all items instead of on only unrecalled items. Thus, in the POK condition, the subjects studied targets, made POK ratings on all items, and then performed the cued recall test. The instructions for the rating phase in the POK condition were identical to those used by Schreiber (1998). The subjects were encouraged to make their ratings as quickly as possible and were told that the best strategy to use was to rely on their initial impression or sense of whether they knew the related study word. FOK ratings were self-paced.

Results and Discussion

Feelings and predictions of knowing. Figure 2 shows the mean of the subjects' median knowing ratings as a function of cue set size, type of knowing rating task, and number of study trials. Two separate ANOVAs were performed on these ratings. In order to facilitate comparison with prior studies using only the standard FOK procedure, the dependent variable in the first analysis included only data for unrecalled items for both FOK and POK conditions. The principal factors in this analysis were cue set size, type of rating task, and number of study trials. The second analysis considered data from the POK task alone, and median ratings for both recalled and unrecalled items were included in the analysis. Thus, the factors in the second analysis were cue set size, number of study trials, and the type of response made in recall (recalled, unrecalled).

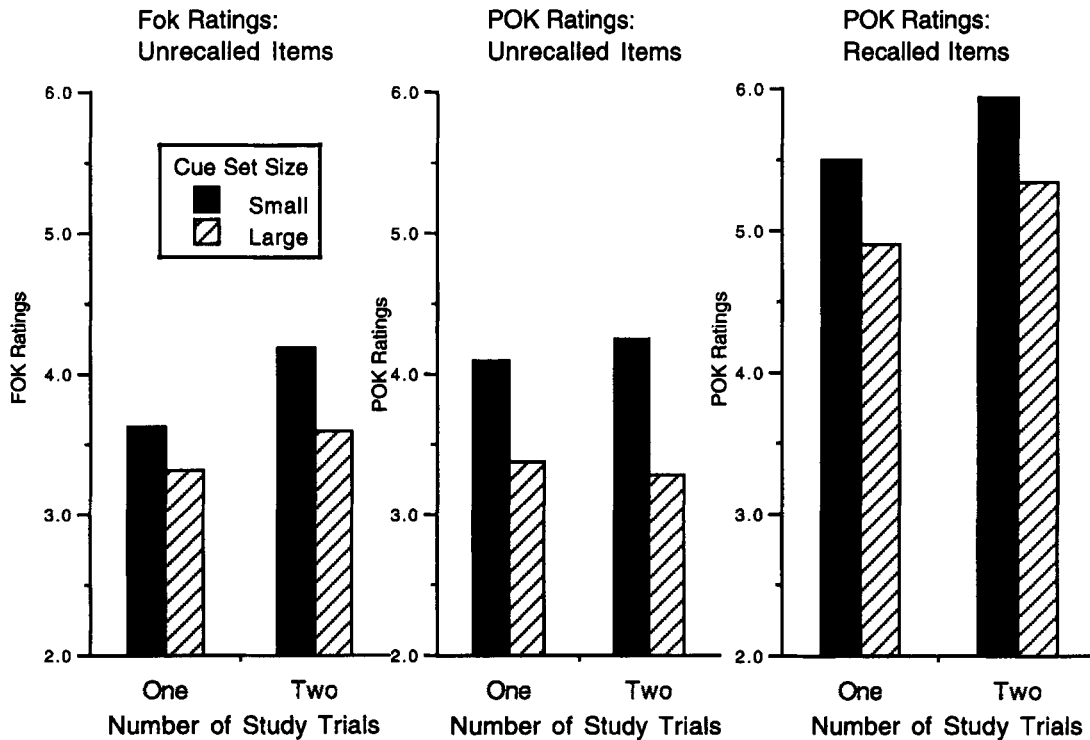


Figure 2. The mean of the individual subjects' median FOK/POK ratings as a function of cue set size and number of study trials in Experiment 2. When using ending stems as the test stimuli, cue set size is inversely related to the magnitude of both FOK and POK ratings, and the same relation holds for both recalled and unrecalled items.

The pattern of findings essentially was identical to the pattern obtained in Experiment 1. In the analysis based on unrecalled items (left and middle panels of Figure 2), ratings were higher for cues defining smaller rhyme categories (4.04) than for cues defining larger rhyme categories (3.39) [$F(1,60) = 33.11$, $MS_e = 0.41$]. Ratings were comparable in the FOK (3.68) and POK (3.75) conditions and were not reliably different ($F < 1$). Finally, ratings were numerically higher following two study trials (3.82) than following one (3.60); however, as shown in Figure 2, this trend was mainly apparent in the FOK condition. Neither the main effect of number of study trials nor the interaction with task was significant (both $F_s < 1$). Again, the most important result was that knowing ratings were inversely related to cue set size.

The rightmost portion of Figure 2 shows the mean of the subjects' median POK ratings for items that were recalled. The data for POK ratings on unrecalled items are the same as in the previous analysis and are shown in the middle panel of Figure 2. With respect to the cue set size factor, similar results were obtained as when only the unrecalled items were considered. Again, ratings were significantly higher for cues linked to smaller sets of rhyming instances (4.94) than for cues linked to larger sets of rhyming instances (4.22) [$F(1,30) = 44.21$, $MS_e = 0.37$]. Ratings also were higher for items that were recalled (5.42) than for those items that were not recalled (3.75) [$F(1,30) = 100.84$, $MS_e = .89$], replicating similar find-

ings reported by Reder and Ritter (1992) and Schreiber (1998). None of the remaining sources of variance were reliable, including the interaction between cue set size and type of recall response and the number of study trials (all $F_s < 1.54$).

Recall. The second and third rows of Table 1 show the mean probabilities of recall for the FOK and POK conditions. Probability of recall was higher for stems defining smaller rhyme categories than for stems defining larger categories and higher for two study trials than for one. The mean probabilities for cues defining smaller and larger categories were, respectively, .66 and .44 [$F(1,60) = 237.43$, $MS_e = 0.01$]. The means for two and one study trials prior to test were, respectively, .61 and .49, and this difference also was reliable [$F(1,60) = 15.82$, $MS_e = 0.03$]. No other factors were reliable sources of variance. These recall findings replicate previous results, now replicated many times (see D. L. Nelson, Schreiber, & McEvoy, 1992, for a review).

The results showed that the magnitude of both POK and FOK ratings was related to the number of rhyming instances linked to the test cues. Ratings for both tasks were significantly higher for ending sounds defining smaller rhyme categories than for sounds defining larger rhyme categories; in the POK condition, the same result was obtained both for recalled and for unrecalled items. This pattern is identical to the pattern found in Experiment 1 in which taxonomic category names were used as

cues and in which knowing ratings were made only on unrecalled items. The results of both experiments are compatible with the inference that knowing ratings are sensitive to the number of different words related to the test cue as measured by the norming procedure, and the results indicate that the same pattern occurs regardless of whether the relation concerns the taxonomic or the rhyme domain or regardless of whether recalled or unrecalled items are involved. Again, these findings appear to be incompatible with the partial-retrieval hypothesis because it assumes that FOK ratings should be lower with fewer neighboring items, and this is the reverse of the pattern that was found. Instead, the findings are more compatible with the hypothesis that knowing ratings are sensitive to the amount of competition between neighboring words that arises when the test cue activates related knowledge.

EXPERIMENT 3

The primary purpose of Experiment 3 was to determine whether an inverse relation between cue set size and FOK and POK ratings could be obtained for associative meaning cues as was the case for taxonomic categories and rhymes. In the study phase, the subjects were shown familiar words (e.g., PADDLE); in the testing phases, they were given meaningfully or associatively related words as cues (e.g., CANOE). In addition to varying cue set size, we again included both the FOK and the POK rating tasks. The main question was whether cue set size influences knowing ratings only for cues in which set membership is defined relatively clearly, as with rhymes and taxonomic categories, or whether this pattern generalizes to associative meaning cues in which set membership is much less well defined. With associative meaning cues, set membership presumably is determined by the many experiences in a person's past in which associations or meaningful relations between words have been established. Once such associations have been established, familiar words presumably activate related words, and competition should be produced. Many recall experiments provide evidence supporting this assumption, because, as with taxonomic category names and ending cues, recall is much more likely, the fewer related associates are activated (reviewed in D. L. Nelson, Schreiber, & McEvoy, 1992). Presumably, if FOK and POK ratings in general are sensitive to competition, both POK and FOK ratings should be higher for associative cues linked to smaller sets.

Method

Design and Subjects. The experimental design was similar to Experiments 1 and 2 except that the subjects received only a single study trial. We used a 2×2 mixed-model design, with cue set size (small, large) manipulated within subjects and with type of knowing rating task (FOK, POK) manipulated between subjects. Fifty students participated in the experiment, with an equal number randomly assigned to each rating task.

Materials. The test cues and targets were constructed with the use of a normative database of approximately 3,500 words (D. L.

Nelson, McEvoy, & Schreiber, 1995). Many groups of participants ($n = 100$ – 200) were given approximately 100 different stimulus words and were asked to write the first word to come to mind that was meaningfully related to or strongly associated with each stimulus word. Their responses were used to estimate both the strength of the preexisting cue-to-target relationship and the associative set size (see D. L. Nelson & Schreiber, 1992, for additional details and rationale). Forty-four cues and 44 targets were selected from the database, representing an equal number of items for each level of cue set size (see Appendix C). In contrast to Experiments 1 and 2, only a single list of cues and targets was used, primarily for practical reasons. The items for Experiment 3 were initially used in a recall experiment. In the recall experiment, two separate lists were constructed, each of which consisted of 22 cue–target pairs. These two lists were combined for the present experiment to ensure a relatively large pool of recalled and unrecalled items. In this list, mean associative set sizes were 7.86 associates ($SD = 0.94$) for cues with smaller sets and 19.18 associates ($SD = 2.61$) for those with larger sets. As before, other attributes of the test cue and target, such as mean preexisting cue-to-target strength, were equated in each condition of cue set size. Mean strengths for these cues were, respectively, .08 ($SD = .05$) and .07 ($SD = .04$), and target set sizes averaged 14.86 associates ($SD = 4.22$) and 15.19 associates ($SD = 4.86$).

Procedure. Pilot testing showed that the intentional learning instructions used in Experiments 1 and 2 produced low levels of recall for the associatively related cues; therefore, in order to obtain higher levels of recall, the subjects were asked to make pleasantness judgments for each target word. After reading each word aloud, the subjects classified it as pleasant, neutral, or unpleasant. The instructions indicated that the purpose of this task was to collect normative data, and subsequent phases of the experiment were not mentioned. The study phase was identical for the subjects in the FOK and the POK conditions, and the target words were presented on a Macintosh Plus computer screen at a 4-sec rate for both conditions.

The general procedures for the FOK and POK conditions were very similar to those used in Experiments 1 and 2, but there were some minor procedural differences related to the use of a computer for the presentation of the stimulus materials. The cued recall instructions indicated that a meaningfully related test cue would be shown for each word in the study list and that the task was to try to recall the related study word. Clarifying examples were provided beforehand. The subjects were asked to try hard to recall as many study words as they could, and they were told that guessing was allowed when they were unsure. Whenever the subjects could not think of the correct study word and did not want to guess, they said the word "Next" and then typed the word "Next" on the keyboard.

In the FOK condition, the FOK phase occurred after recall responses were given to all of the test cues, as before. Also as before, the instructions for FOK ratings indicated that only the cues for unrecalled items would be presented. The computer was programmed to present cues only for unrecalled items, which were presented one at a time in a random order. On each trial, the subjects were asked to read each cue silently and to indicate their FOK for the study word using the same 6-point scale as before. As soon as the subjects decided on a rating for an item, they said the rating aloud. After giving their rating verbally, the subjects typed it on a numeric keypad located on the right portion of the keyboard. Similar procedures were used in the POK condition, except the subjects assigned to this condition completed the knowing rating phase prior to the cued recall test. In both conditions, the subjects were given practice items to familiarize them with the rating procedure.

Results and Discussion

Feelings and predictions of knowing. The left and middle portions of Figure 3 show knowing ratings for

unrecalled items as a function of cue set size and rating task. As with Experiments 1 and 2, ratings were higher the fewer related items were linked to the test cue. Averaged across subjects, median ratings for cues with smaller and larger associative sets were, respectively, 3.53 and 3.00 [$F(1,48) = 5.95, MS_e = 0.81$]. Ratings were numerically higher in the FOK condition than in the POK condition, but this difference was not reliable [$F(1,48) = 2.75, MS_e = 3.49$]. The interaction between cue set size and type of knowing rating also was not reliable ($F < 1$). The pattern of findings was similar when both recalled and unrecalled items in the POK condition alone were analyzed (see middle and right portions of Figure 3). Ratings were higher for cues with smaller associative sets (4.56) than for cues having larger associative sets (4.14).³ In addition, ratings were higher for items that were eventually recalled (5.73) than for those that were not recalled (2.97) (see Reder & Ritter, 1992, Schreiber, 1998, and Experiment 2 of the present series for similar findings). The main effects of cue set size [$F(1,24) = 7.58, MS_e = 0.58$] and type of recall response [$F(1,24) = 80.63, MS_e = 2.36$] were reliable, but the interaction between these sources was not ($F < 1$), indicating that the same inverse relation between knowing ratings and cue set size was apparent regardless of whether the target could be recalled.

Recall. Mean probabilities of recall are shown in the last two rows of Table 1. Probability of recall was higher for cues linked to smaller sets of associates (.68) than for

those linked to larger sets of associates (.54), and recall was more likely in the FOK condition (.65) than in the POK condition (.57). Both the effect of cue set size [$F(1,48) = 80.12, MS_e = 0.01$] and the effect of rating task were reliable [$F(1,48) = 6.88, MS_e = 0.03$], and the interaction between these sources was not [$F(1,48) = 1.95$]. We discuss the implications of these findings more fully in the General Discussion section.

GENERAL DISCUSSION

The findings of all three experiments were consistent in showing that both FOK and POK ratings were inversely related to the number of different concepts linked to test cues as measured by normative data. Both types of ratings invariably were higher when test cues were linked to smaller sets of related items than when they were linked to larger sets of related items. Similar findings were obtained for taxonomic category names, ending sounds, and meaningfully related words. The fact that similar findings were obtained for three very different kinds of cues indicates that such findings are robust and may generalize to retrieval cues of many types. Given that FOK and POK ratings were inversely, rather than directly, related to cue set size, the present findings appear to be at odds with one assumption made by the partial-retrieval hypothesis (Blake, 1973; Eysenck, 1979; Koriat, 1993; Koriat & Lieblich, 1977). This hypothesis assumes that FOKs should increase as more information is associated with a test cue. The more neighboring items activated in memory, the higher the FOK. This assumption suggests that FOKs should be lower for cues linked to smaller sets than for those linked to larger sets. Thus, a central assumption currently included in the partial-retrieval hypothesis that seems to be directly relevant to cue set size predicted a pattern of results opposite to the pattern that was obtained.

Although the inverse relation between cue set size and FOKs appears to be inconsistent with one key assumption of the partial-retrieval hypothesis, it is important to note that the present findings do not suggest that FOKs never are based on the retrieval of partial information. It is quite possible that, under some circumstances, FOKs may be based on the deliberate retrieval of partial information, such that the more information accessed, the more positive the sense of knowing. In the case when people have a TOT experience, they apparently can retrieve related information, and this information could be used as a basis for computing FOKs. In addition, the results of several studies (e.g., Koriat, 1993; Krinsky & T. O. Nelson, 1985) indicate that when people experience a retrieval failure and are inclined to report related information (i.e., commission errors), they have more positive FOKs than when they have a retrieval failure and are not inclined to report any information at all (i.e., omission errors). Both kinds of findings suggest that in at least some cases there is a positive correlation between FOKs and the amount of information that is retrieved overtly.

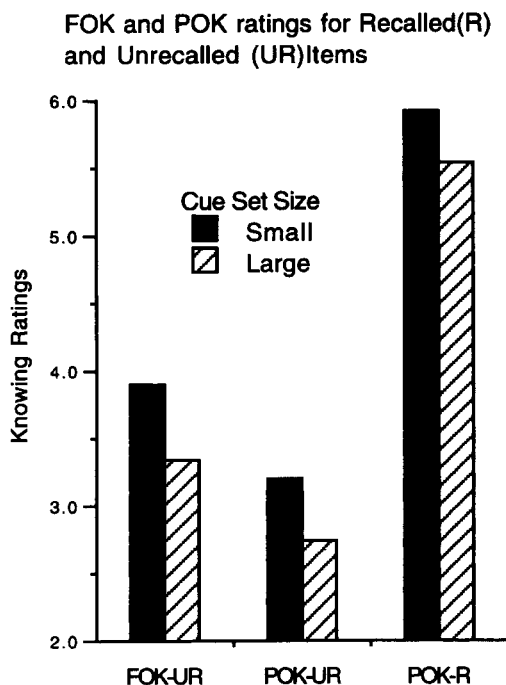


Figure 3. The mean of the individual subjects' median FOK/POK ratings as a function of cue set size in Experiment 3. When using associative meaning cues, cue set size is inversely related to the magnitude of both FOK and POK ratings, as was the case for taxonomic categories and rhymes.

Table 2
The Relation Between FOK and POK Ratings,
Cue Set Size, and Type of Recall Error in Experiment 3

Recall Error	FOK Ratings		POK Ratings		<i>M</i>
	Small Cue Set	Large Cue Set	Small Cue Set	Large Cue Set	
Omission	3.97	3.47	2.83	2.41	3.17
Commission	4.47	4.00	3.76	3.13	3.84
<i>M</i>	4.22	3.73	3.28	2.77	

Note—FOK, feeling of knowing; POK, prediction of knowing.

In fact, some aspects of the findings we obtained are consistent with the partial-retrieval assumption. For example, Table 2 shows median FOK and POK ratings, averaged across subjects, as a function of type of recall error made and cue set size for Experiment 3. Note that knowing ratings always were higher for commission error responses than for omission error responses. This result is consistent with prior research and is compatible with the partial-retrieval view. However, even more important for present purposes, note also that Table 2 shows very clearly that an equivalent inverse relation between the magnitude of knowing ratings and cue set size was obtained regardless of the type of recall error made. Averaged across subjects, median knowing ratings always were higher for cues linked to smaller associative sets than for those linked to larger sets, regardless of whether the analysis included only items that led to errors of commission (4.22 vs. 3.73) [$t(36) = 2.17$] or whether it included only items that led to errors of omission (3.28 vs. 2.77) [$t(44) = 3.00$].⁴ In other words, even with type of recall error controlled, there still was an inverse relation between knowing ratings and cue set size. Those data provide a nice contrast because they include findings that are handled easily by the partial-retrieval assumption and findings that seem to require additional assumptions. Put simply, it appears that additional assumptions would need to be made in order for the partial-retrieval view to account for the equivalent cue set size pattern that was obtained regardless of type of recall error. One assumption we consider in more detail later on is that cue set size might not be associated primarily with the tendency to retrieve and to report related information overtly. Rather, it may be more associated with the manner in which preexisting knowledge relevant to performance in some task is organized and activated implicitly in long-term memory. The bottom line is that the present findings suggest that the partial-retrieval hypothesis as currently articulated may not apply to all situations because FOKs were inversely, rather than directly, related to the number of items linked to the test cue, and this same relation was evident for various different types of cues, for both POK and for FOK ratings, for recalled and unrecalled items, and for omission and for commission responses.

In the remainder of this article, we consider several possible ways to reconcile the present findings with the partial-retrieval hypothesis, and we then discuss the relevance of the findings for other hypotheses. One poten-

tial way to reconcile the present findings with the partial-retrieval hypothesis would be to add the assumption that people sometimes base their FOKs on strategic estimates of set size. People might conceivably use the strategy of estimating how much competing information is linked to a cue and then make the reasonable inference that the probability of recalling a recently studied event should be higher when the test cue is related to a small number of potential answers, relative to when it is related to a larger number. Although some aspects of the present findings are compatible with this explanation, there are several reasons why this kind of strategy-based explanation becomes strained in trying to account for the entire range of set size findings.

In a series of experiments that conceptually is very relevant to the present investigation, Schreiber (1998) found that, with cue set size and a large variety of other attributes of the test cue and target equated, both FOK and POK ratings were higher when the studied target word had fewer preexisting links to related items. An inverse relation between meaning-related target set size and knowing ratings was obtained even when rhyme cues were given as the test stimulus. Thus, in Schreiber's (1998) investigation of target set size, the same pattern of findings occurred as in the present investigation involving cue set size. It is difficult to imagine how people successfully could use any type of set size estimation strategy when the number of associates of the target word rather than the test cue is varied, especially when rhyme cues are given as the initiating stimulus. Presumably, with rhyme cues, people consciously must focus on information from an entirely different domain of information than the meaning-related associates of the target, yet meaning-related target set size was associated with performance in the same way as cue set size.

A second reason why a set size estimation strategy likely would not be able to account for the entire range of set size findings is based on additional findings indicating that an inverse relation between cue set size and FOKs can be obtained given stimuli for which people appear to have difficulty making accurate estimates of set size. Schreiber (1993) compared the correlations between strategic estimates of set size for the three types of cues that were used in the present experiments and set size based on association norms. One group of participants ($n = 30$) was presented with the taxonomic categories from Experiment 1, and another group ($n = 30$) was presented with the ending stems from Experiment 2. Both groups were

asked to estimate category size using a rating scale appropriately labeled for each type of category. Correlations between these ratings and category size as measured by the appropriate normative databases were computed in order to determine estimation accuracy. The mean correlation was $+0.43$ for the group given taxonomic categories, and it was $+0.27$ for the group involving ending stems. These correlations provide support for the hypothesis that people have some ability to estimate the sizes of taxonomic categories and a lesser ability to estimate the sizes of rhyme categories. The magnitude of these correlations agrees relatively well with results obtained by Joelson and Herrmann (1978), who used taxonomic category names. They found a correlation of $+0.25$ between estimates of category size and category size measured using norms. These correlations indicate that it is possible that the subjects participating in Experiments 1 and 2 could have based their knowing ratings on estimates of category size with a moderate to low degree of success. Schreiber (1993) presented a third group of subjects with the associative cues that were used in Experiment 3. Using a similar rating scale, the subjects were asked to estimate how many other words are closely related in meaning or strongly associated with each stimulus word. The expectation was that correlations between estimation ratings and associative set size measured using the norms would be much lower than for the other types of cues because, in contrast to rhymes and taxonomic categories, associative sets inherently are defined less clearly and because category size information is not likely to be stored directly for associates. This expectation was fulfilled; the correlation for associative cues was only $+0.02$ ($n = 35$).

A low correlation between associative set size and strategic estimates of associative set size was expected for several reasons. Assume that, in trying to estimate the sizes of associative sets, people use the reasonable strategy of generating related words explicitly and then counting how many were generated. Such a strategy is likely to lead to misperceptions of associative set size because of *response chaining* and *retrieval inhibition*. To illustrate these problems, assume a person is given a cue word such as TIGER and generates STRIPES, ZEBRA, HORSE, DONKEY, and so on. One problem is that the further along in this chain, the less the generated words are directly related to the stimulus that initiated the process (e.g., consider TIGER and DONKEY), and a second, related problem is that the more such words are generated, the more they inhibit or block people from generating other words that are directly related to the initial stimulus. Retrieval inhibition and response chaining are well-known effects in studies examining the generation of information from semantic memory (Freedman & Loftus, 1971; Nickerson, 1984; Roediger & Neely, 1982), and both factors could easily lead to misperceptions of the number of associated concepts that are activated initially when a familiar word is encoded, which presumably is what is measured using the association norms in question. In fact, we intentionally use a single-response technique in collecting the normative data, rather than a multiple-response technique,

precisely to minimize such problems (cf. Libukman, 1994; McEvoy & D. L. Nelson, 1982; D. L. Nelson & Schreiber, 1992; D. L. Nelson, Schreiber, & McEvoy, 1992). For rhyme and taxonomic categories, the effects of chaining and inhibition may be less than those of associates because generated items that clearly do not belong to the category in question can be edited out and suppressed more easily. In contrast to taxonomic and rhyme categories, associative sets are less well defined, and this may magnify response chaining and inhibition effects. Finally, it is unlikely that category size information is stored directly for associates. In contrast, such information probably is more likely to be stored for at least some taxonomic categories (e.g., "What are the *nine* planets of the solar system?"). For present purposes, the most important observation is that, for whatever reason, subjects apparently have difficulty strategically estimating associative set size accurately. At the same time, association norms are quite accurate predictors of performance in recall and a variety of other tasks (e.g., priming tasks), and measurements of both strength and set size produce very consistent and reliable patterns of results (D. L. Nelson & Schreiber, 1992; D. L. Nelson, Schreiber, & McEvoy, 1992). Given these observations, the question that remains is, why did cue set size influence both POK and FOK ratings in Experiment 3 in which associative meaning cues were employed? Both the finding that there is only a very weak correlation between associative set size and subjects' strategic estimates of associative set size and the important findings that manipulations of target set size and cue set size for various kinds of stimuli produce identical patterns of results seem to point to a different kind of explanation.

Another, and, we believe, more fruitful, way to try to reconcile the present findings with the partial-retrieval hypothesis would be to begin by making a distinction between more automatic and more strategic uses of information. Presumably, any task, including FOK and POK ratings, is influenced by some combination of both kinds of information. Set size effects may be more closely associated with relatively automatic activation processes, whereas the conscious and deliberate retrieval of related information may be more closely associated with strategic processes. The literature on set size in various tasks supports this assumption. This literature indicates that cue and target set size are inversely related to performance on implicit and explicit memory tests (D. L. Nelson, Bajo, & Canas, 1987; D. L. Nelson et al., 1988; D. L. Nelson, Schreiber, & Holley, 1992). Such findings do not depend on the use of particular strategic orientations toward the test cue and target during encoding or retrieval (D. L. Nelson & McEvoy, 1979; D. L. Nelson, Schreiber, & Holley, 1992); they are obtained when the stimulus word is shown at extremely brief exposure durations of 200 msec or less (Schreiber & Carter, 1998), and meaning-related target set size influences performance even when the subjects' attention is focused on sound rather than meaning during the encoding and retrieval trial (D. L. Nelson, Bajo, & Canas, 1987; D. L. Nelson & Friedrich, 1980;

D. L. Nelson, Schreiber, & Holley, 1992; Schreiber, 1998). The latter finding is noteworthy because it suggests that preexisting memories in the meaning domain can become activated and cause competition-like effects regardless of a person's degree of conscious attention to meaning during either the study or the test trial. On the basis of what we know, it appears that set size findings are associated with relatively automatic processes.

The hypothesis that FOKs can be influenced by relatively automatic processes is also consistent with other findings. Drawing again on prior work involving target set size, Schreiber (1998) found that both FOK and POK ratings were higher when the studied target word was linked to fewer items. In addition, the same inverse relation between target set size and knowing ratings was obtained when subjects used different encoding strategies during the study phase, such as rating target words for pleasantness or naming their vowels. These encoding strategies produced large differences in recall performance, and they even influenced judgments of knowing. Most important for present purposes, however, are the findings that, regardless of the particular encoding strategy used, knowing ratings always were higher when targets were linked to smaller associative sets. The finding of an inverse relation between target set size and FOKs, even when the subjects were simply required to name vowels of the studied word, is particularly important because it is compatible with the inference that focusing attention on relatively low level perceptual features of letters contained in target words is sufficient to activate a word's associates and to produce competition that can be detected later on when making knowing ratings. Deliberate and conscious attention to meaning apparently is not required to activate a word's associates and to influence subsequent judgments of knowing. Schreiber (1998) also found that meaning-related target set size influenced knowing ratings when subjects named vowels of the target during the study phase, and, during the knowing rating phase, they were given ending sounds as the initiating stimulus. As such, the number of meaning-related associates of the target affected knowing ratings despite discouraging subjects from consciously thinking about meaning both during the study phase and during the testing phase. It would appear to be difficult to explain these findings without assuming that knowing ratings can be based, in part, on the relatively automatic activation of closely related concepts.

The activation hypothesis essentially assumes that people sometimes are conscious of only the end product of the process used to compute FOKs, rather than the process itself. With respect to the cue set size findings, the end product provides a sense of the degree to which information is known, which in turn is based, in part, on the amount of competition that arises when a test cue is encoded. Subsequent to this initial encoding, a person very well could choose to retrieve related information, and it is this act that appears to be captured particularly well by the partial-retrieval hypothesis. However, it is this act that also may not be the primary source of the cue set size findings. Such findings may be tied less to the subsequent strategic retrieval process and more to the process

by which closely related items are activated in long-term memory initially. Clearly, it is possible for the activation and partial-retrieval hypotheses to coexist if their assumptions are not construed as being mutually exclusive.

The findings showing that cue set size was related to FOK and POK ratings also have implications for another explanation of FOK effects. The cue familiarity hypothesis (e.g., Metcalfe, 1993; Reder, 1988; Reder & Ritter, 1992; Schwartz & Metcalfe, 1992) assumes that FOK and POK ratings are influenced by a cue recognition process. The more prior experiences a person has had with a test cue, the more likely it will seem familiar and be recognized and the higher the FOK. This hypothesis seems relevant to the present findings because it makes assumptions about the role played by information linked to the test cue, and it may thus potentially provide an explanation for the cue set size findings. However, there are several reasons why a cue recognition process is unlikely to be successful. First, it is important to note that the test cues in the present experiments were not available either before or during the study trial; therefore, they could not have acquired differential familiarity during the experimental session. Second, there is nothing in the research literature suggesting that words with smaller sets inherently are perceived to be more familiar. In general, set size is only very weakly correlated with other variables that could influence familiarity and, hence, recognition. D. L. Nelson and Schreiber (1992) examined the relationship between associative set size and a variety of other variables. On the basis of a sample of 2,172 words, they found correlations of $-.03$ with printed frequency, $.16$ with log of printed frequency, and $-.14$ with word concreteness. In addition, for 56 taxonomic categories, Joelson and Herrmann (1978) found correlations of $.14$ with log of printed frequency and $-.01$ with concreteness. Libukman (1994) investigated the rhyme set sizes of 926 words and found correlations of $.06$ with printed frequency and $.09$ with concreteness. Finally, and perhaps most important, despite many attempts to find set size effects in standard recognition tasks, the number of associates linked to a word appear to have no effect on its recognition (e.g., Canas & D. L. Nelson, 1986; D. L. Nelson, Canas, & Bajo, 1987; D. L. Nelson, Canas, Casanueva, & Castano, 1985; D. L. Nelson et al., 1988). Prefamiliarizing subjects with the words to be used as retrieval cues is known to affect judgments of knowing, and this important finding has now been replicated several times (e.g., Metcalfe, 1993; Reder, 1988; Schwartz & Metcalfe, 1992). As such, it seems reasonable to conclude that cue recognition plays an important role in such judgments. However, it seems that some mechanism other than cue recognition would be needed to explain the present findings because set size does not influence recognition and essentially is unrelated to variables that are known to influence recognition. A cue recognition mechanism also would appear to have even more difficulty explaining target set size findings (Schreiber, 1998).

A relatively straightforward way to explain the inverse relation between cue set size and FOKs assumes that

competition is produced when the test cue activates connections to neighboring concepts. Although, the activation of such connections does not appear to influence familiarity or recognition, it can produce competition in a variety of tasks (D. L. Nelson, Schreiber, & McEvoy, 1992). When a cue is presented during testing, connections to closely related neighboring concepts presumably are activated relatively automatically. The activation of such connections provides immediate and rapid access to preexisting knowledge linked to a given stimulus. At the same time, what is activated produces competition in the form of background noise in the same way that background noise is produced in a room filled with many people who are talking simultaneously. The more noise, the greater the amount of interference produced. When there is less noise in the form of fewer activated connections to related concepts, there is less competition with activation received by the target. As a consequence, FOKs are higher when the cue activates connections to fewer neighboring concepts than when it activates connections to a greater number of neighboring concepts. These relatively simple assumptions can be used to explain the inverse relation between cue set size and FOKs for various types of cues; with the addition of other similar assumptions, it also would be possible to explain related findings involving target set size under different conditions of encoding and testing (see Schreiber, 1998).

REFERENCES

- BLAKE, M. (1973). Prediction of recognition when recall fails: Exploring the feeling of knowing phenomenon. *Journal of Verbal Learning & Verbal Behavior*, *6*, 89-94.
- BROWN, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin*, *109*, 204-223.
- CANAS, J. J., & NELSON, D. L. (1986). Recognition and environmental context: The effect of testing by phone. *Bulletin of the Psychonomic Society*, *24*, 407-409.
- EYSENCK, M. W. (1979). The feeling of knowing a word's meaning. *British Journal of Psychology*, *70*, 243-251.
- FREEDMAN, J. L., & LOFTUS, E. F. (1971). Retrieval of words from long-term memory. *Journal of Verbal Learning & Verbal Behavior*, *10*, 107-115.
- GILLUND, G., & SHIFFRIN, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 1-67.
- HART, D. (1967). Memory and the memory-monitoring process. *Journal of Verbal Learning & Verbal Behavior*, *6*, 685-691.
- JOELSON, J. M., & HERRMANN, D. J. (1978). Properties of categories in semantic memory. *American Journal of Psychology*, *91*, 101-114.
- KINTSCH, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 163-182.
- KORIAT, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, *100*, 609-639.
- KORIAT, A., & LIEBLICH, I. (1977). A study of memory pointers. *Acta Psychologica*, *58*, 151-164.
- KRINSKY, R., & NELSON, T. O. (1985). The feeling of knowing for different kinds of retrieval failure. *Acta Psychologica*, *58*, 141-158.
- LEONESIO, R. J., & NELSON, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *16*, 464-470.
- LIBUKMAN, T. M. (1994). Norms for words that rhyme. *Behavior Research Methods, Instruments, & Computers*, *26*, 278-322.
- McEVOY, C. L., & NELSON, D. L. (1982). Category name and instance norms for 106 categories of various sizes. *American Journal of Psychology*, *95*, 581-634.
- METCALFE, J. (1993). Novelty monitoring, metacognition, and control in a composite holographic associative recall model: Implications for Korsakoff amnesia. *Psychological Review*, *99*, 1-20.
- NELSON, D. L. (1989). Implicitly activated knowledge and memory. In C. Izawa (Ed.), *Current issues in cognitive psychology: The Tulane Floweree Symposium on Cognition* (pp. 129-162). Hillsdale, NJ: Erlbaum.
- NELSON, D. L., BAJO, M. T., & CANAS, J. J. (1987). Prior knowledge and memory: The episodic encoding of implicitly activated associates and rhymes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *13*, 54-63.
- NELSON, D. L., CANAS, J. [J.], & BAJO, M. T. (1987). The effects of natural category size on memory for episodic encodings. *Memory & Cognition*, *15*, 133-140.
- NELSON, D. L., CANAS, J. [J.], CASANUEVA, D., & CASTANO, D. (1985). Prior knowledge and recognition. *American Journal of Psychology*, *98*, 379-397.
- NELSON, D. L., FRIEDRICH, M. A. (1980). Encoding and cueing sounds and senses. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *6*, 717-731.
- NELSON, D. L., & McEVOY, C. L. (1979). Encoding context and set size. *Journal of Experimental Psychology: Human Learning & Memory*, *5*, 292-314.
- NELSON, D. L., McEVOY, C. L., & BAJO, M. T. (1988). Lexical and semantic search in cued recall, fragment completion, perceptual identification, and recognition. *American Journal of Psychology*, *101*, 465-480.
- NELSON, D. L., McEVOY, C. L., & SCHREIBER, T. A. (1995). *The University of South Florida word association, rhyme, and word fragment norms*. Unpublished manuscript.
- NELSON, D. L., & SCHREIBER, T. A. (1992). Word concreteness and word structure as independent determinants of recall. *Journal of Memory & Language*, *31*, 237-260.
- NELSON, D. L., SCHREIBER, T. A., & HOLLEY, P. E. (1992). The retrieval of controlled and automatic aspects of meaning on direct and indirect tests. *Memory & Cognition*, *20*, 671-684.
- NELSON, D. L., SCHREIBER, T. A., & McEVOY, C. L. (1992). Processing implicit and explicit representations. *Psychological Review*, *99*, 322-348.
- 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.
- NELSON, T. O., & NARENS, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125-141). San Diego: Academic Press.
- NICKERSON, R. S. (1984). Retrieval inhibition from part-set cuing: A persisting enigma in memory research. *Memory & Cognition*, *12*, 531-552.
- RAAIJMAKERS, J. G. W., & SHIFFRIN, A. M. (1983). Search of associative memory. *Psychological Review*, *88*, 93-134.
- REDER, L. M. (1988). Strategic control of retrieval strategies. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 227-259). New York: Academic Press.
- 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.
- ROEDIGER, H. L., III, & NEELY, J. H. (1982). Retrieval blocks in episodic and semantic memory. *Canadian Journal of Psychology*, *36*, 213-242.
- SCHACTER, D. L. (1983). Feeling of knowing in episodic memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *9*, 39-54.
- SCHREIBER, T. A. (1993). *Feelings of knowing are based upon implicitly activated knowledge about the test cue and the target*. Unpublished doctoral dissertation, University of South Florida.
- SCHREIBER, T. A. (1998). The effects of target set size on feelings of knowing and cued recall: Implications for the cue effectiveness and partial-retrieval hypotheses. *Memory & Cognition*, *26*, 553-571.
- SCHREIBER, T. A., & CARTER, K. (1998). *The activation of preexisting associations in an episodic memory task*. Manuscript submitted for publication.
- SCHREIBER, T. A., SERGENT, S. D. (1998). *Dissociate effects of cue set*

size on predictions of knowing and actual knowing. Manuscript submitted for publication.

SCHWARTZ, B. L., & METCALFE, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 1074-1083.

NOTES

1. The acronym stands for *processing implicit and explicit representations*.

2. We thank Thomas O. Nelson, a reviewer of this article, for the suggestion to report medians instead of means. The primary advantage of using medians is that one must assume only that subjects are using at least an ordinal scale in making their ratings, rather than interval or ratio scales.

3. We have replicated the FOK and POK results from all three experiments a number of additional times using similar sets of stimuli e.g., Schreiber & Sergent, 1998.

4. Scores for the subjects who did not make any omission or commission errors in one of the conditions were omitted from that particular statistical analysis.

**APPENDIX A
Materials for Experiment 1**

Cue	Target	
	List 1	List 2
Small Cue Set Size		
STINGING INSECT	MOSQUITO	WASP
PRECIOUS GEM	EMERALD	RUBY
KIND OF JEWELRY	WATCH	BRACELET
ARMED SERVICES	MARINES	NAVY
BIRD OF PREY	EAGLE	VULTURE
BRASS INSTRUMENT	TUBA	TROMBONE
DAILY MEAL	SUPPER	BREAKFAST
PART OF A FACE	MOUTH	CHEEK
HAIR COLOR	BRUNETTE	AUBURN
PART OF SPEECH	PRONOUN	ADJECTIVE
SINGING VOICE	BARITONE	TENOR
US. GOVERNMENT BRANCH	JUDICIAL	LEGISLATIVE
REFERENCE BOOK	ENCYCLOPEDIA	THESAURUS
PART OF AN ATOM	NEUTRON	ELECTRON
COMPASS DIRECTION	SOUTH	WEST
AMERICAN COIN	PENNY	DIME
UNIT OF LENGTH	YARD	METER
ROOM IN A HOUSE	BATHROOM	KITCHEN
KIND OF SEASON	WINTER	SUMMER
TWO-WHEELED VEHICLE	BICYCLE	MOTORCYCLE
EATING UTENSIL	SPOON	KNIFE
Large Cue Set Size		
TYPE OF CLOTH	WOOL	SILK
DOG BREED	COLLIE	SHEPHERD
TYPE OF FASTENER	ZIPPER	BUTTON
FELONY CRIME	ROBBERY	ARSON
TYPE OF FLOWER	TULIP	DAISY
TYPE OF FRUIT	PEAR	GRAPE
KIND OF LIQUOR	WHISKEY	VODKA
MEDICAL SPECIALTY	SURGEON	NEUROLOGIST
TYPE OF NUT	ACORN	PECAN
TYPE OF SCIENCE	PHYSICS	ASTRONOMY
PROFESSIONAL SPORT	TENNIS	HOCKEY
WILD ANIMAL	TIGER	BEAR
PART OF A BOAT	BOW	STERN
CHEMICAL ELEMENT	HYDROGEN	SODIUM
ARTICLE OF CLOTHING	BLOUSE	DRESS
TYPE OF COSMETIC	ROUGE	MASCARA
MYTHICAL BEING	CYCLOPS	CENTAUR
TYPE OF SHIP	FREIGHTER	YACHT
KIND OF WOOD	PINE	OAK
TYPE OF BUILDING	SKYSCRAPER	HOUSE
TYPE OF CANDY	CHOCOLATE	PEPPERMINT

APPENDIX B
Materials for Experiment 2

Small Cue Set Size		Large Cue Set Size	
Cue	Target	Cue	Target
List 1			
AINT	SAINT	AB	CRAB
ALF	HALF	ACK	TRACK
ANNEL	FLANNEL	AIL	MAIL
ARGE	CHARGE	AIN	MAIN
ARM	CHARM	AM	JAM
AST	LAST	AP	MAP
AVEL	TRAVEL	ART	QUART
EM	STEM	ASH	CASH
ERVE	NERVE	ATE	PLATE
ESH	FRESH	AVE	GRAVE
IEF	THIEF	EEK	WEEK
IFE	WIFE	ELL	SHELL
IRST	THIRST	EST	WEST
ISK	BRISK	ET	WET
OIN	JOIN	ICE	PRICE
OLT	VOLT	INE	PINE
OOD	WOOD	INK	THINK
OOK	SPOOK	OAL	COAL
ORK	STORK	OCK	ROCK
ORM	STORM	ONE	BONE
OST	FROST	OP	STOP
OTH	CLOTH	ORE	STORE
OUSE	SPOUSE	OSE	ROSE
OVE	GLOVE	OUP	SOUP
OWN	CROWN	OW	SNOW
ULK	HULK	UE	GLUE
UNCH	PUNCH	UG	HUG
UTCH	CRUTCH	UM	DRUM
List 2			
ALT	SALT	ADE	TRADE
ANCH	BRANCH	AG	FLAG
ARD	YARD	AKE	SNAKE
ARP	SHARP	ANK	BANK
ATH	PATH	ARE	CARE
ATTLE	RATTLE	ASH	RASH
EARCH	SEARCH	ASTE	PASTE
EAST	LEAST	AY	PAY
EG	KEG	EAM	DREAM
ENCH	QUENCH	EAN	CLEAN
ERM	TERM	EAT	HEAT
ETCH	STRETCH	EEL	WHEEL
IFT	SWIFT	EEZE	FREEZE
ISH	WISH	END	FRIEND
ISP	CRISP	ICK	BRICK
ISS	BLISS	IDE	WIDE
IVER	RIVER	ILL	HILL
OBE	ROBE	IN	WIN
OOF	HOOF	ING	KING
ORCE	FORCE	JP	SHIP
ORTH	WORTH	IST	WRIST
OST	POST	IZE	SIZE
OTCH	SCOTCH	OAD	ROAD
OD	PROUD	OB	KNOB
OVE	GROVE	OKE	JOKE
UCE	SPRUCE	OOM	BROOM
URE	CURE	ESS	GUESS
USK	HUSK	UMP	JUMP

APPENDIX C
Materials for Experiment 3

Small Cue Set Size		Large Cue Set Size	
Cue	Target	Cue	Target
SIGHT	VISION	FRIEND	COMPANION
STEM	LEAF	FIRE	FLAME
SURGEON	OPERATION	FACT	TRUTH
TROPHY	AWARD	DOOR	LOCK
SCENT	PERFUME	EGG	YOLK
ZODIAC	ASTROLOGY	CRIMINAL	JUSTICE
REPTILE	FROG	CRISP	CRUNCH
RECTANGLE	TRIANGLE	DAMP	DRY
LAUNDRY	WASH	CARD	ACE
MONK	MONASTERY	GHOST	GOBLIN
GEM	RUBY	GRAPE	RAISIN
INFANT	CHILD	GRIP	HAND
ELK	ANTLER	HARP	VIOLIN
DIME	COIN	TOUR	TRAVEL
AUTHOR	POET	STRAW	HAY
BANQUET	FEAST	SYSTEM	PROCESS
BARRACUDA	SHARK	TOMATO	SOUP
BUCKLE	STRAP	TRUCK	VAN
CANOE	PADDLE	TURN	SPIN
SIMPLE	COMPLEX	ABILITY	SKILL
SOLDIER	WARRIOR	TWINE	ROPE
ROOF	FLOOR	MONTH	WEEK

(Manuscript received December 9, 1996;
revision accepted for publication July 28, 1997.)