Feeling of knowing and duration of unsuccessful memory search

MURRAY SINGER AND HEATHER L. TIEDE
University of Manitoba, Winnipeg, Manitoba, Canada

Five experiments examined the impact of feeling of knowing on decisions to continue or to terminate the search of memory in question answering. First, two pairs of experiments respectively scrutinized knowledge about (1) ordinary facts and (2) national capitals. The first experiment of each pair extracted normative data: The participants indicated whether they had probably once known the answer to a question (*once-knew-it* scale), supplied the answer if they knew it, and either judged the likelihood of their recognizing the answer or made other pertinent metacognitive judgments. In these norming experiments, recognition ratings were highly correlated with once-knew-it responses, and both measures were highly predictive of performance. This indicated that both measures reflect feeling of knowing judgments. In the second experiment of each pair, different participants were timed as they indicated whether they knew the answer to the same questions. Response latencies for responding "don't know" were strongly positively correlated with the once-knew-it judgments made by the norming participants. This relationship was corroborated by Experiment 5, which compared the crucial measures *within* participants. These outcomes suggest that, in this context, feeling of knowing judgments are predictive of how long people will search memory for requested information.

What comes to mind when you encounter the question "Who composed the opera *Carmen*?" What if the answer does not readily pop into your head? This could occur whether or not you were an opera fan. If you were not, you might quickly decide and report that you do not know who composed *Carmen*. As a knowledgeable opera fan, in contrast, you might wrack your brain at length in an attempt to retrieve the answer, but you still might fail to do so.[1]

The goal of this study was to examine the quality of people's unsuccessful search of memory for requested information. In this regard, the typical North American might be equally unsuccessful in identifying the capitals of Ghana and Australia. However, Australia's relative familiarity might result in a lengthy search, even if that search proved to be unsuccessful.

A number of early studies suggested that people can often quickly decide that they do not know something. Lindsay and Norman (1977) observed that one ought to be able to rapidly indicate that one does not know "Hemingway's phone number." In an empirical study, Glucksberg and McCloskey's (1981) participants learned facts such as *Bill has a rifle* and subsequently had to judge the truth of related facts. For a test item such as *Bill has a pencil*, the participants were instructed to answer "don't know," following the rationale that Bill might or might not also have had a pencil. Correct "don't know" responses were made faster than "true" and "false" ones were. Glucksberg and McCloskey proposed that the participants could answer "don't know" as soon as it was determined that there was

no connection between the two crucial concepts, like *Bill* and *pencil*.

Similarly, Singer's (1981, 1984) participants read such sentences as "*the pilot painted with the brush*" and were expected to respond "don't know" to test sentences such as "*the pilot painted a fence.*" As predicted, mean correct "don't know" latencies were faster than "no" latencies. Singer's response criteria differed from those of Glucksberg and McCloskey (1981), but his explanation was similar: He proposed that participants responded "don't know" as soon as they discovered that the antecedent sentence included no semantic case information (Fillmore, 1968) relevant to the focus of the test item.

It is clear, however, that not all "don't know" decisions are fast ones. Kolers and Palef (1976) asked their participants to decide whether they had ever visited various cities. The "no" responses are of greatest relevance to the present concerns. Participants were fastest to report that they had never visited Asian cities, not as fast for European cities, and slowest for North American cities. This suggested that some aspect of a city's familiarity influenced the duration of the memory search. Likewise, Glucksberg and McCloskey (1981, Experiment 3) found that people quickly answer "don't know" to general-knowledge facts unsupported by relevant information (e.g., "*Ann Landers has a law degree*") but slowly to facts for which relevant information is available ("*Ann Landers has a journalism degree*").

One striking feature of Glucksberg and McCloskey's (1981) findings was that correct "don't know" response

M. Singer, m_singer@umanitoba.ca

times (RTs) to "*Bill has a pencil*" were greater if the participant had explicitly learned that it was unknown whether Bill had a pencil (an "explicit don't-know") than if no connection were made between "Bill" and "pencil." Furthermore, Klin, Guzmán, and Levine (1997) demonstrated that the relative magnitude of RTs for explicit and implicit "don't know" items could be manipulated by regulating the relatedness of a question, with reference to a prose passage. In particular, higher relatedness was associated with longer "don't know" replies.

The present study scrutinized people's decisions to either continue or terminate memory search, with particular emphasis on the time needed to report not knowing the answer to general-knowledge questions. This issue is widely considered to reflect people's feelings of knowing an answer (FOKs; e.g., Hart, 1965). The remainder of the introduction will provide an overview of the FOK framework, describe previous findings of the relationship between FOK and "don't know" RT, and then outline the present approach.

Feelings of Knowing for General-Knowledge Questions

Metacognitive judgments of knowing a response or being able to later recognize a response are referred to as FOK judgments, an essential aspect of memory self-monitoring insofar as such judgments bear directly on decisions to initiate, sustain, and eventually terminate a search of long-term memory. In a standard FOK experiment, participants are presented with either (1) novel stimuli representing to-be-remembered information (e.g., paired associates) or (2) general-knowledge questions (e.g., "What is the capital of Ghana?"). FOK judgments are frequently elicited only after recall failure by asking participants the likelihood of their recognizing or later recalling the correct answer (but see Koriat, 1993). People's FOK ratings are predictive of later performance in tasks ranging from the recognition of general-knowledge answers (Nelson & Narens, 1980) and newly learned facts (Hart, 1965) to problem solution (Metcalfe, 1986).

Metacognitive assessments of knowing are widely held to be distinct from, and preliminary to, processes of memory retrieval (Klin et al., 1997; Reder, 1987; Singer, 1990). FOKs are considered to be automatically, and therefore quickly, available (Reder & Schunn, 1996). They reflect the familiarity of the question (e.g., Koriat & Levy-Sadot, 2001; Nelson, Gerler, & Narens, 1984; Reder, 1987), the retrievability of full or partial responses (Koriat, 1993; Koriat & Levy-Sadot, 2001), and/or one's metamemorial knowledge about the question (e.g., Costermans, Lories, & Ansay, 1992). Consistent with the proposed automaticity of FOK assessments is that people need less time to make judgments about the likely success of memory search than to report the corresponding answer (Koriat & Levy-Sadot, 2001; Reder, 1987).

Furthermore, quickly available FOKs can influence whether one applies a memory-retrieval strategy or a computational strategy to a test probe. Reder (1982) presented evidence that, in the evaluation of test items with reference to text—for example, "Was the race held on a cold day?"—

people attempted to directly retrieve recently encountered (hence, apparently familiar) items, but to judge the plausibility of delayed items—for example, "Well, the race was in July, so it probably wasn't cold." Likewise, Reder and Ritter (1992) documented a preference for retrieving the answers to arithmetic problems (e.g., 17×23) that seemed familiar, because one or both of their operands had occurred in the corresponding positions of previously encountered problems. Computation, in contrast, was preferred for unfamiliar problems (see also Reder, 1987; Young, 2004).

Feeling of Knowing and Latency of "Don't Know" Responses

There is converging evidence of a distinct positive correlation between "don't know" RTs (DKRTs) and FOK. However, the present study used a methodology somewhat different from these antecedent investigations. To highlight the capacity of varying methods to clarify the FOK–DKRT relationship, we next describe (1) prior investigations of this issue; (2) an overview of the present approach; and (3) the merits of our methodology.

Previous investigations. Prior studies bearing on the FOK–DKRT relationship have utilized either (1) the standard FOK procedure or (2) experimental manipulations of FOK. Using the standard method, researchers have typically scrutinized the relationship between FOK and memory performance, with emphasis on general-knowledge facts (Costermans et al., 1992; Nelson et al., 1984; Nelson & Narens, 1980). The methods of quantifying FOK have varied slightly, encompassing 5-point (Costermans et al., 1992) and 9-point (Nelson & Narens, 1980) rating scales and participants' ranking of individual items (Nelson et al., 1984). FOK estimation has usually but not always (Costermans et al., 1992) been contingent on recall failure. Recognition has been the typical memory criterion task (e.g., Costermans et al., 1992), but Nelson et al. (1984) inspected perceptual identification and multiple-choice retrieval. Nelson and Narens (1980) used no criterion task, because they were simply norming a sample of general-knowledge questions. These studies were consistent in their detection of positive correlations between FOK and the time to report not knowing an answer, with values of $r \approx .50$ regularly measured (Costermans et al., 1992; Nelson & Narens, 1980).

In *experimental manipulations of FOK*, FOK has been posited to be enhanced by (1) the familiarity of a question cue and the accessibility of information pertinent to the question (Koriat & Levy-Sadot, 2001); (2) the familiarity of question contents and the degree of elaboration of answer-bearing text (Klin et al., 1997); and (3) the frequency of appearance of top and bottom operands in arithmetic problems (Reder & Ritter, 1992). The dependent variables were derived from participants' answers to questions pertinent to the stimuli. For example, Klin et al. presented text-related questions that could be answered "yes," "no," or "don't know." It is noteworthy that some of these studies have tacitly (or explicitly; Schreiber, 1998) solicited an assessment of FOK, rather than the more usual request to report fact retrieval. Thus, Koriat and Levy-Sadot asked their participants to indicate, without

answering the test item, whether or not they wished their answer to count in a subsequent game-show simulation (see also Reder, 1987). Likewise, Reder and Ritter's task required participants to signal whether they intended to answer an arithmetic problem by retrieving or computing the result. Presumably, both Koriat and Levy-Sadot's "don't count" preference and Reder and Ritter's "compute" decision reflected low FOK assessments.

Like the standard studies, these experimental investigations uniformly revealed that "don't know" RTs are greater in conditions associated, according to the scientific hypothesis, with high FOK as compared with low FOK. For example, Klin et al. (1997, Experiment 3) detected greater DKRTs for high-familiarity questions than for low-familiarity ones. Finally, it is noteworthy that DKRT increases systematically with FOK, for both general-knowledge (Koriat & Levy-Sadot, 2001; Reder & Ritter, 1992) and newly learned (Klin et al., 1997) facts.

Overview of the Present Study

We adopted the approach of obtaining general-knowledge-fact FOK ratings (e.g., "the largest desert on earth") from one sample of participants in norming studies and examining different participants' latencies to report not knowing these facts. Several features of our method were particularly useful in the characterization of the FOK–DKRT relationship. We next consider those features.

Deriving FOK and criterion measures from different samples. Researchers have commonly obtained FOK ratings and criterion measures from the same individuals to determine whether FOK is predictive of successful memory retrieval (e.g., Hart, 1965; Nelson et al., 1984). One drawback of this approach is that the first of two judgments might affect the second one. In assessing the FOK–DKRT relationship, for example, the participant's memory of a high FOK rating might promote longer search, and vice versa. Either of these circumstances could inflate the apparent correlation between FOK and DKRT.

A complementary approach has been to obtain FOK norms and criterion measures from different participant samples. In this regard, FOK is systematically related to (1) the normed familiarity of general-knowledge facts (Koriat & Levy-Sadot, 2001, Experiment 2A; Nelson, Leonesio, Landwehr, & Narens, 1986) and (2) objective characteristics of taxonomic categories (Schreiber, 1998; Young, 2004). Performance generally correlates more closely with one's own FOK (Nelson et al., 1986) and judgment of learning (Hertzog, Kidder, Powell-Moman, & Dunlosky, 2002; King, Zechmeister, & Shaughnessy, 1980) than with the metacognitive judgments of others. This trend indicates that people have privileged access to their own experiences with a stimulus (Hertzog et al., 2002; Nelson et al., 1986). Nevertheless, performance is usually, though not always (Nelson et al., 1986), correlated with others' knowledge norms.

Scales of feeling of knowing. The participant's FOK has typically been measured on multipoint discrete scales (e.g., Costermans et al., 1992; Hart, 1965; Nelson & Narens, 1980). These metrics have ordinal properties (Nelson, 1984) and have served their intended purpose effectively.

In contrast, the experimental studies bearing on the FOK–DKRT relationship, discussed earlier, were based on dichotomous rather than multipoint treatments of FOK. In this regard, FOK has been operationalized in terms of the manipulation of fact familiarity and fact accessibility (Klin et al., 1997; Koriat & Levy-Sadot, 2001). These studies frequently, although not always (Koriat & Levy-Sadot, 2001), have yielded the predicted significant effects with regard to the DKRT dependent variable. However, they have not provided high-resolution assessments of the FOK–DKRT relationship.

Our procedure assessed FOK using both traditional scales and novel dichotomous classifications of whether participants thought they had once known the answer. Because our FOK norms and our timed criterion measure were obtained from different participant samples, the analytic focus was on items rather than participants: We measured correlations between FOK and DKRT *means*, derived from different samples. These scales were continuous rather than dichotomous, enabling a richer characterization of the relationships among our variables (e.g., Hertzog et al., 2002).

Reports of retrieval versus reports of feeling of knowing. We noted earlier that some experimental investigations that have revealed a positive relationship between FOK and DKRT measured the time needed to generate an FOK assessment, rather than to report retrieval (Koriat & Levy-Sadot, 2001; Reder & Ritter, 1992). However, because metacognitive evaluations can be dissociated from retrieval processes, those results are suggestive rather than definite with regard to the focus of the present investigation. Our criterion measures (Experiments 2, 4, and 5) all concerned reports of retrieval rather than assessments of knowing.

Contingent and noncontingent FOK ratings. In the original methodology, FOK ratings were solicited only for those items that the participant did not recall (e.g., Hart, 1965). According to this approach, it was only for failed items that scrutiny of the participant's ability to predict performance was useful. However, investigators have increasingly scrutinized metacognitive judgments not contingent on retrieval failure (e.g., Costermans et al., 1992; Koriat, 1993; Koriat & Levy-Sadot, 2001; see also Schreiber's, 1998, prediction-of-knowing ratings). This procedure yields richer datasets. For example, disregarding successful items is likely to exclude those items associated with the highest degrees of FOK. Success on an item need not imply very high FOK, as in the instance of a correct guess.

In within-samples studies of FOK, measuring an item's FOK and criterion for a given participant can be either contingent or noncontingent. However, the between-samples FOK procedure is inherently associated with noncontingency: The norming sample simply provides a rating for every item. Experiments 1–4 adopted the between-samples approach; therefore, FOK measurement was noncontingent. In conclusion, we are proposing not that contingent measurement of FOK prohibits informative investigations of these issues, but that noncontingent measurement yields the richest possible dataset.

Conclusion. In summary, there is considerable evidence that people can make fast "don't know" responses when they can readily determine that the stimulus is unfamiliar, or that no relevant information is available. However, this relationship has been the direct focus of few studies. There are growing indications that "don't know" decisions are pertinent to a considerable variety of tasks, ranging from signal detection to eyewitness testimony (Scoboria, 2004). In view of these considerations, the present study was undertaken to measure DKRT for general-knowledge questions across the full range of FOKs.

EXPERIMENT 1

The main purpose of Experiment 1 was the collection of knowledge assessment norms concerning general-knowledge facts. A second goal, however, was the evaluation of a particular metacognitive judgment considered to be specifically relevant to the registration of a "don't know" reply. Therefore, two sets of norms were collected. For the first, the participants were asked to indicate whether they had probably *once known* a fact ("once knew it," or OKI), and then to supply the answer if possible (cf. Costermans et al., 1992). In the second norming procedure, the participants provided the more traditional rating: whether they thought they would recognize the answer (Hart, 1965; Nelson et al., 1984; Nelson & Narens, 1980). Thus, Experiment 1 provided the desired norms and determined whether the OKI and probability-of-recognition ratings reflect similar metacognitive processes.

Method

Participants. The participants were 69 native-English-speaking male and female introductory-psychology students at the University of Manitoba. The participants received course credit for their participation. The OKI procedure was administered to 34 participants and the recognition rating procedure to 35 participants.

Materials. A priority of the study was to identify facts that covered, in a relatively uniform fashion, the full range of the OKI scale. Norms were collected for 76 facts. Many of them were derived from Nelson and Narens's (1980) fact recall norms. First, 5 facts were randomly chosen from each 5% interval of Nelson and Narens, ranging from 0.0%–5.0% to 14.1%–20.0%. Then, another 5 facts were randomly selected from each of Nelson and Narens's 10% ranges of 20.1%–30.0% to 80.1%–90.0%. To these 55 facts were added 21 facts that had very high recall-probability scores in a pilot study. Some of the latter facts were from Nelson and Narens's norms and others were constructed for the present study. Several of the new facts had Canadian content.

The facts were worded as descriptive phrases such as "largest desert on earth," rather than explicit questions such as "What is the largest desert on earth?" This choice was made because we wanted to use identical phrases in the norming and answer-latency experiments; and the shorter, descriptive phrases were likely to minimize variability among the Experiment 2 RTs.

For both procedures, the test materials consisted of booklets containing an instruction page and the list of 76 test items. About half of the participants in each procedure received the items in a randomly determined "forward" order, and the other half received a reversed "backward" order.

Procedure. Two norming procedures were used. One participant group was asked to indicate by a "yes" or "no" reply whether they "had probably ever known" the answer to each of 76 general-knowledge prompts. After replying, the participant supplied the

answer, if known. In the second procedure, different participants rated each item on a 9-point scale ranging from 1 (*sure I wouldn't recognize the answer*) to 9 (*sure I would recognize the answer*). Participants also supplied the answer, if known.

The data were collected in two group sessions. In each session, the participants were randomly assigned to receive one of the two norming procedures. They read the instructions and performed their task. Each session lasted 30 min.

Results and Discussion

OKI procedure. All statistical procedures in this study used the criterion of $\alpha = .05$. The descriptive statistics for OKI, recognition rating, and accuracy are shown in Table 1. Both OKI and recognition rating were negatively skewed: For example, of the 76 facts, 22 had OKI scores between 89% and 94%.

The accuracy values shown in Table 1 stem from the OKI procedure. In that procedure, the mean accuracies were 38.1% and 39.0% for the forward and backward forms of the questionnaires, respectively. The correlation of accuracy across the two forms was $r = .95$. In view of the high correlation of accuracy between the two forms of the OKI booklet, accuracy was not scored in the recognition booklets.

The correlation coefficients among the measures of Experiment 1 appear in Table 2. All of these values were highly significant [$rs(74) \geq .72$]. It is noteworthy that OKI was almost perfectly correlated with recognition rating ($r = .95$), which has typically been used as an index of FOK. This suggests that the two measures reflect much the same metacognitive knowledge.

EXPERIMENT 2

Experiment 2 was designed to clarify the relationship between people's DKRTs for general-knowledge facts and their feeling of knowing those facts.

Method

Participants. The participants were 32 new individuals from the pool that was accessed in Experiment 1. Due to a programming error, the data of 3 participants were lost.

Materials. The experimental materials were 54 facts selected from the 76 evaluated in Experiment 1. In view of the very high correlation between OKI and recognition rating in Experiment 1, we decided to focus on the OKI scores. First, all of the facts with OKI scores smaller than 88% and accuracy scores greater than 2% were selected, yielding 20 facts. Second, from facts with OKI scores of 88% or greater, ones were selected that were as uniformly distributed in accuracy as possible. Third, from facts with Experiment 1 accuracy scores of 2% or lower, ones were selected that were as uniformly distributed in OKI as possible. In the test list, the questions were phrased descriptively, using the exact wording that had been employed in the norm booklets of Experiment 1.

Table 1
Descriptive Statistics of Experiments 1 and 2

Measure	M	SD	Skew
Once knew it	66.4	29.1	−0.7
Recognition rating	6.8	2.0	−0.7
Accuracy	38.6	34.2	0.3
DKRT (msec)	4,126	820	−0.2

Note—DKRT, "don't know" response time.

Table 2
Correlations Among the Measures of Experiments 1 and 2

Measure	OKI	RR	Acc	DKRT
Once knew it (OKI)		.95	.72	.50
Recognition rating (RR)			.79	.50
Accuracy (Acc)				.24

Note—DKRT, "don't know" response time.

Procedure. The sessions were conducted with groups of 1 to 4 individuals. The stimuli were displayed on monochrome monitors at a distance of 40 cm. Each group encountered the 54 facts in a different random order. On each trial, a fixation point appeared for 1 sec, followed immediately by a phrase describing a fact. The participants used "yes" and "no" keys to indicate whether the answer was known. There was an answer time limit of 8 sec, and the participants were instructed to answer as quickly as possible. During a subsequent 8-sec intertrial interval (ITI), the participant wrote either the answer "yes" if the "yes" key had been pressed, or "no" if the "no" key had been pressed. Singer (1986, Experiments 1 and 2) showed that the response-key task yields a pattern of latencies similar to that derived from a vocal onset task. The buttonpress response has been used fruitfully in the study of decisions of not knowing (Glucksberg & McCloskey, 1981) and FOK (Koriat & Levy-Sadot, 2001, Experiment 2A).

Results

The descriptive statistics and the correlations for the DKRTs appear in Tables 1 and 2, respectively. DKRT correlated $r = .50$ with both OKI and recognition rating (ps $< .01$). The correlation between DKRT and accuracy was only marginally significant ($r = .24$, $p = .08$).

In view of the uncertain psychometric properties of the measures, Kendall's tau, a nonparametric correlation statistic, was computed for all pairs of variables (Kendall & Stuart, 1967; Nelson, 1984). Tau yielded a pattern of significant correlations identical to r, except that the correlation between DKRT and accuracy reached full significance ($\tau = .26$, $p < .01$).

Because the correlation values were almost identical for the two norming scales, the relationship between DKRT and the norms was pursued with reference to the OKI scores only. Figure 1 shows the scattergram relating DKRT and OKI for the 54 items of Experiment 2. The best-fitting regression line relating the two variables had the formula RT = $(3,375 + 14 \times \text{OKI})$ msec. In the regression analysis, it was
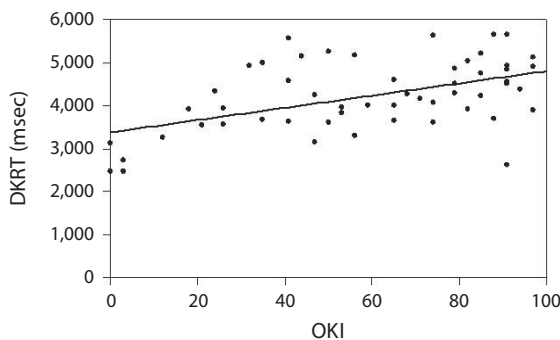

Figure 1. Scattergram relating the "don't know" response times (DKRTs) of the 54 items of Experiment 2 to the once-knew-it (OKI) norms of Experiment 1.

taken into consideration that the test phrases varied from 4 to 16 syllables. However, DKRT was not significantly correlated with number of syllables ($r = .12$). Therefore, syllable length was not used as a regression predictor variable.

Although the focus of this study was the "don't know" RTs, it was noted that the mean latency of "yes" answers for the 54 stimuli of Experiment 2 was 3,659 msec ($SD = 1,007$ msec). The correlation between the "yes" latencies and the OKI scale was not significant ($r = .11$).

Discussion

Several features of Experiment 2 were noteworthy. First and most importantly, we detected a strong monotonic relationship between metacognitive assessments of knowing and the duration of search prior to the registration of a "don't know" reply. Although there have been prior reports of such a relationship, some previous studies have treated FOK as a dichotomous rather than a continuous variable (Klin et al., 1997; Koriat & Levy-Sadot, 2001).

Second, people's judgments of having once known a fact (OKI) were very strongly correlated with ratings of anticipated recognition, a more traditional index of FOK. The OKI judgment bears particular relevance to "don't know" judgments: The evaluation that one has never known a fact would seem to provide a sound basis for curtailing memory search. The success of the OKI measure is consistent with the evidence of Costermans et al. (1992) that metacognitive judgments of the participant's prior experience are central to FOK.

Unlike the standard FOK approach (Hart, 1965), the metacognitive measures of Experiment 1 were not contingent on retrieval failure, and were obtained from a participant sample distinct from the one that generated the performance measure. It was noted in the introduction that both of these features are increasingly common ones in FOK research (Hertzog et al., 2002; King et al., 1980; Koriat & Levy-Sadot, 2001; Nelson et al., 1986; Schreiber, 1998; Young, 2004). Noncontingent FOK ratings have the merit of reliably extracting the desired information for all items rather than only for a subset of them. We propose that using separate FOK and performance samples makes sense, on the plausible assumption that, for a vast domain of general knowledge, the state of knowledge of one sample of individuals is predictive of the memory performance of a different sample.

To summarize, Experiments 1 and 2 confirmed a strong correlation between FOK, in the form of the OKI measure, and DKRT. In addition, these experiments revealed that people's judgments of having known a fact act as an index of FOK. OKI ratings derived from one sample of participants acted as an excellent predictor of another sample's DKRTs. The next two experiments examined this relationship in a different context.

EXPERIMENT 3

General-knowledge questions are highly variable in their length, grammatical construction, and content. To more fully evaluate the relationship between FOK and DKRT, we conducted an additional pair of experiments,

which focused on a homogeneous set of facts with well-defined answers. The chosen domain was that of national capitals. Most countries have a single capital with a one- or two-word name. Most adult North Americans have studied geography and have likely memorized numerous national capitals. This realm of facts was expected to offer a broad range of instances of people's actual knowledge, as well as of their estimation of their familiarity with these facts. Another advantage of national capitals is that in a timed answering experiment, it is possible to express a question like "What is the capital of Ecuador?" in the single word *Ecuador*. This would reduce variability of reading time that might tend to obscure the relationship between FOK and "don't know" RT.

Method

In Experiment 3, norms generally analogous to those of Experiment 1 were collected for 60 national capitals. Twelve countries each were chosen from Europe, East Asia, the Middle East (including North Africa), sub-Saharan Africa, and South America. The selection was not made randomly, and the only restriction was that countries with multiword names were excluded.

One aim of the norming study was to identify countries whose capitals varied considerably in familiarity, but it was recognized that familiarity within the five regions would not show similar variation (Friedman, Brown, & McGaffey, 2002; Kolers & Palef, 1976). For example, the selected European countries seemed highly familiar, whereas the African countries were quite unfamiliar.

The test booklets consisted of an instruction page and two pages of country names. The participants answered up to three questions for each country. They first indicated whether they had heard of the country. If they had, they next indicated whether or not they "probably once knew the capital." If they again responded affirmatively, they finally wrote the name of the capital, if they could. For convenience, the resulting scales will be referred to as the *never-heard-of-country* scale, the *once-knew-it* (the capital) norms, and accuracy.

Two forms of the booklet were prepared. In the first, the country names were presented in a random order. In the second, the reverse order was used.

The data were collected in two group-testing sessions with 52 naive participants from the same population as before. In each session, the two forms of the booklet were received by about half of the participants. The participants read the instructions and completed the booklet at their own pace. The task took about 30 min.

Results and Discussion

The descriptive statistics of Experiment 3 appear in Table 3. The mean OKI value was 46.0%. These scores ranged from 2% for several countries to 96% for Italy. Participants reported never having heard of a country 16.0% of the time. This measure ranged from 0% for 23 countries to 85% for Gabon. As a result, never-heard-of-country was positively skewed. Finally, the mean accuracy of identification of capitals was 9.0%.

Table 4 shows that the three measures of knowledge were strongly and significantly correlated ($df = 58$) with one another in the expected ways. The OKI-versus-accuracy correlation indicates that if a large proportion of participants indicated that they had probably once known a capital, the corresponding accuracy score was relatively high. High never-heard-of-country scores were associated with low OKI and accuracy scores. Kendall's tau revealed an identical profile of significant outcomes.

Table 3
Descriptive Statistics of Experiments 3 and 4

Measure	M	SD	Skew
Once knew it	46.0	30.1	0.1
Never heard of country	16.0	25.3	1.7
Accuracy	9.0	15.2	2.4
DKRT (msec)	2,523	618	0.5

Note—DKRT, "don't know" response time.

One merit of the OKI norms is that the countries were quite uniformly distributed along the scale. In contrast, both the accuracy and the never-heard-of-country scores were severely skewed. Therefore, it would be harder to assess the relationship between DKRTs and either of the other two measures. Accordingly, the particular goal of Experiment 4 was to scrutinize the relationship between people's latencies to report not knowing the capital of a country and the corresponding OKI scores of Experiment 3. The other knowledge-estimation norms of Experiment 3 were also considered.

EXPERIMENT 4

Method

The stimuli were 9 country names randomly selected from each of the 5 regional sets of 12 inspected in the Experiment 3 norms, for a total of 45. The sessions were conducted with groups of up to 4 individuals. Each group viewed the list of 45 country names, with a different random order being used for each group. The participants pressed response keys labeled "yes" and "no" to indicate whether they knew the capital of the country. The procedure was identical to that of Experiment 2 except that, because of the single-word stimuli, the time limit for answering was 5 sec rather than 8 sec. The participants were 30 individuals from the same pool as before, who had not participated in Experiments 1–3. A technical error resulted in the loss of the data of 1 participant.

Results

"No" responses ("I don't know the capital") were registered 74.1% of the time, with a mean latency of 2,523 msec ($SD = 618$ msec). The correlations between the "no" latencies (i.e., the DKRTs) and the norming measures of Experiment 3 appear in Table 4. The OKI-versus-DKRT correlation was $r = .84$. The DKRTs were also strongly positively correlated with Experiment 3 accuracy and strongly negatively correlated with the Experiment 3 never-heard-of-country scale. The latter value means that never having heard of a country was associated with a very brief memory search for the capital.

Figure 2 shows the scattergram relating DKRT to the OKI norms for the 45 items of Experiment 4. The graph portrays a highly systematic relationship. The correspond-

Table 4
Correlations Among the Measures of Experiments 3 and 4

Measure	OKI	NHC	Acc	DKRT
Once knew it (OKI)		−.76	.70	.84
Never heard of country (NHC)			−.35	−.61
Accuracy (Acc)				.67

Note—DKRT, "don't know" response time.

ing regression line is described by the formula RT = $(1{,}721 + 17 \times \text{OKI})$ msec.

There were 7 countries for which no participant responded "yes" to signal knowing the capital. For the remaining 38 countries, the mean "yes" RT was 2,640 msec ($SD = 745$ msec). The correlation between OKI and the "yes" RT was not significant ($r = -.21$).

Discussion

Experiments 3 and 4 again assessed the relationship between FOK and "don't know" answer latencies. The national-capital stimuli were uniform in content, had well-defined answers, and could be expressed succinctly. They covered the full range of the OKI index of FOK. The main result, depicted in Figure 2, was that DKRT increased monotonically with OKI. Consistent with our assumptions about the impact of the uniformity of content and stimulus length, the DKRT-versus-OKI correlation (.84) was significantly larger than the value of .50 detected in Experiment 2 ($z = 3.20$, $SE = 0.21$). Experiments 3 and 4 also offered corroboration that people's judgments of having once known a fact (OKI) serve as a useful index of FOK. OKI was again strongly correlated with the norming participants' (Experiment 3) accuracy in addition to the DKRTs.

In view of the focus on national capitals, the norms of Experiment 3 included the metacognitive assessment of whether the participant had ever heard of the country in question. This measure showed a sensible negative correlation of $r = -.61$ with the DKRTs. However, the absolute value of that correlation was significantly smaller than that between OKI and the DKRTs ($z = 2.34$, $SE = 0.22$).

In view of the use of answer time limits, it is noteworthy that the mean RT in each of Experiments 2 and 4 was approximately half of its limit. Therefore, it appears that the participants were able to scale their RTs, arguably precisely, to the time available (e.g., Ratcliff & McKoon, 1989).

In neither Experiment 2 nor Experiment 4 were the "yes" RTs significantly correlated with OKI. Prior evidence concerning this relationship is somewhat sparse. In a study of dichotomously manipulated FOK, Koriat and Levy-Sadot (2001) examined the impact of both fact familiarity and fact accessibility on the latency of feeling of knowing judgments. For example, the topic "Who is the current president of Yale University?" is familiar, but the answer to the question is of low accessibility. Koriat and Levy-Sadot reported that "yes" replies were faster for high- than for low-FOK facts, in terms of both familiarity and accessibility. In Nelson and Narens's (1980) examination of people's world knowledge, the researchers detected what they considered to be a modest negative correlation ($r = -.30$) between correct RTs and FOK.

The absence of a significant correlation between OKI and the "yes" RTs in the present study may be due, in part, to the small number of relevant observations. The mean "yes" reply rates were 48.9% and 25.9% in Experiments 2 and 4, respectively. This study was designed to maximize the opportunity to evaluate the relationship between FOK and "don't know" RTs. It would take a different design to permit closer scrutiny of the relationship between FOK and affirmative latencies.

EXPERIMENT 5

Experiments 1–4 suggested that, in the domains of general knowledge and national capitals, FOK judgments of one sample are predictive of the performance of a different participant sample. These outcomes are consistent with our assumption about the relative uniformity of knowledge across individuals from the same population. Moreover, our procedure permitted us to obtain FOKs that were not contingent upon retrieval failure.

The goal of Experiment 5 was to extend these findings to a within-participants comparison of DKRTs and FOK judgments about general-knowledge facts. Demonstrating such a relationship would strengthen our claim that, in the context of answering general-knowledge questions, higher feelings of knowing an answer are associated with longer times to respond "don't know."

Method

Participants. The participants were 44 naive students from the same population as in Experiments 1–4. The data of 2 participants were lost due to a programming error. The RT data of another participant were excluded due to spurious responding (latencies under 200 msec).

Materials. First, two norming procedures were used in order to select 60 novel facts from a pool of 115 items. Two naive groups of individuals from our usual participant population provided either recognition ratings on a 9-point scale ($n = 62$) or OKI judgments ($n = 61$). These measures were elicited using the wording of Experiment 1. However, the items were presented in random order on a computer monitor, rather than in a test booklet. No accuracy information was collected. The 54 experimental and 6 filler items chosen for Experiment 5 received recognition ratings ranging from 2.5 to 8.0 and had OKI scores ranging from 8% to 85%. As in Experiment 1, facts were selected that were uniformly distributed in OKI and recognition ratings. The facts were expressed in short, descriptive phrases.

Procedure. The experimental sessions were conducted with groups of 1–4 individuals. The stimuli were displayed on computer
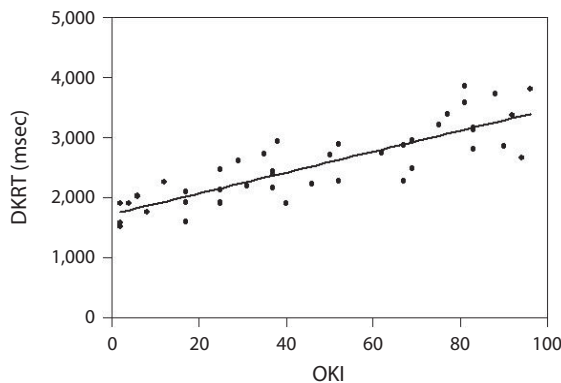

Figure 2. Scattergram relating the "don't know" response times (DKRTs) of the 45 items of Experiment 4 to the once-knew-it (OKI) norms of Experiment 3.

monitors. FOK ratings and question-answering times were measured in successive experimental phases. Distinct instructions preceded each phase. In Phase 1, FOK judgments were elicited using the 9-point recognition-rating method of Experiment 5. In Phase 2, the participants indicated whether they knew the answer about each fact and then, if possible, supplied the answer. The 54 experimental stimuli were presented in random order. The filler facts comprised the first and last 3 items in the list.

On each trial of Phase 1, a left-adjusted fixation point was presented for 500 msec, followed immediately by a general-knowledge fact. The participants indicated on a 9-point scale their likelihood of later recognizing the response. The rating scale was displayed above the facts to make unnecessary people's reliance on their memory of the scale. A 6-sec RT limit was imposed in order to elicit relatively fast FOK judgments. There was a 1-sec ITI.

In Phase 2, the items were presented on the screen, generally following the procedure of Experiment 2. The participants used keys labeled "yes" and "no" to signal whether they knew the answer. There was a 5-sec RT limit. The participants then had 8 sec during which to write either the answer, if they had responded "yes," or "no" if they had so signaled. Then, upon pressing their space bars, the participants encountered the next item.

Results and Discussion

The participants failed to reply within the time limits on 2.2% of Phase 1 trials (FOK) and 10.4% of Phase 2 trials (timed judgment of knowing). Table 5 shows the descriptive statistics for the remaining trials. Across items, the recognition ratings ranged from 2.38 ("scientific study of grasses") to 7.95 ("Manitoba's current premier").

The Phase 2 "no" response rate was 65.7% ($SD =$ 20.8%), with an average RT of 2,219 msec ($SD =$ 441 msec). The correlations between FOK (the Phase 1 recognition ratings) and answer times were computed as follows. In Phase 2, each participant answered "no" to a subset of the 54 experimental items, and "yes" to the complementary subset. Therefore, for each participant, it was possible to calculate the correlation between FOK and DKRT for the former subset and between FOK and "yes" RT for the latter subset. In view of the impact of syllable length of the stimulus questions upon RT, all correlations partialled out the effect of syllable length.

The mean within-participants partial correlation between FOK and DKRT was .40, a value that differed significantly from zero [$t(40) = 10.68$]. The mean within-participants partial correlation between FOK and "yes" RTs was −.32. This value likewise differed significantly from zero [$t(40) = −5.68$]. The appreciable correlation between FOK and DKRT further corroborates the findings of Experiments 2 and 4.

Experiment 5 focused on the within-participants relationship between FOK and DKRT. For comparability with Experiments 2 and 4, however, we also computed the

corresponding within-items correlations, again partialling out the effect of syllable length. As in the other experiments, FOK and DKRT were positively correlated ($r =$ .62). In contrast, the partial correlation between FOK and "yes" RT was $r = .00$.

GENERAL DISCUSSION

There is extensive evidence that people's metacognitive judgments of knowing can predict their memory performance (Hart, 1965; Nelson & Narens, 1980) and regulate the strategies of memory search (Reder, 1987; Reder & Ritter, 1992; Young, 2004). One domain of memory search that has received increasing attention during the past 30 years is people's decisions of not knowing something (Glucksberg & McCloskey, 1981; Klin et al., 1997; Kolers & Palef, 1976; Scoboria, 2004; Singer, 1984). Several regularities have emerged from this research. When the content or relationship expressed in a test fact is highly unfamiliar, reports of not knowing are fast (Glucksberg & McCloskey, 1981; Singer, 1986). However, the duration of search preceding the "don't know" report increases systematically with familiarity (Costermans et al., 1992; Nelson & Narens, 1980). Indeed, the latencies both of unsuccessful memory searches and of metacognitive judgments reflecting the ignorance of a fact increase systematically with experimentally enhanced familiarity (Klin et al., 1997; Koriat & Levy-Sadot, 2001; Reder & Ritter, 1992).

In this study, our participants' metacognitive judgments about once having known a fact (OKI) were very strongly correlated with the more traditional recognition-rating index of FOK (Experiment 1). Consistent with this observation, OKI was strongly correlated throughout with the time people took to search memory before reporting that they did not know the answer to general-knowledge questions.

Following the present rationale, one might likewise anticipate that high FOK would be associated with low latencies of reporting that an answer *was* known. Evidence consistent with this hypothesis was considered in the Discussion of Experiment 4 (Koriat & Levy-Sadot, 2001; Nelson & Narens, 1980). Using our between-samples methodology, we detected no systematic relationship in this regard. However, when, in Experiment 5, the focus was on judgments within individuals, FOK and "yes" response latencies exhibited the predicted negative correlation. This outcome is consistent with the previously discussed tenet of FOK—namely, that people have privileged access to their own experiences with a stimulus (Hertzog et al., 2002; Nelson et al., 1986). Accordingly, assessing the association between FOK and other variables *within* individuals, as in Experiment 5, maximizes the opportunity of detecting systematic relationships. Conversely, the nonsignificant FOK-versus-"yes" RT correlations in Experiments 2 and 4 may have resulted from the relatively small number of "yes" responses or the low power associated with between-participants designs.

The methods of Experiments 1–4 differed from those of the standard FOK procedure (Hart, 1965; Nelson & Narens, 1980) in (1) our use of distinct samples of par-

Table 5
Descriptive Statistics of Experiment 5

Measure	M	SD
Phase 1 recognition rating (FOK)	5.4	1.2
Phase 1 accuracy (%)	26.5	13.6
DKRT (msec)	2,219	441
"Yes" RTs (msec)	2,394	543

Note—FOK, feeling of knowing; DKRT, "don't know" response time.

ticipants to generate the norming data and the judgment latencies and (2) the evaluation, by the norming participants, of every stimulus, as opposed to the subset of test stimuli that result in retrieval failure. These issues were addressed in the discussion of Experiment 2.

During the past 20 years, the preponderance of FOK research has been directed toward understanding the processes underlying FOK judgments and their accuracy (e.g., Koriat, 1993; Reder, 1987). In contrast to these investigations, little attention has been directed at the means by which FOK judgments are obtained. Predictions concerning the likelihood of future remembering are customarily measured using recognition-rating scales, percentage estimates of later recognition, identification, or recall of the answer. Of particular importance is that FOKs are typically generated by asking participants the likelihood of future remembering based on some intuitive perception of knowing a response. In this sense, FOKs represent not only an assessment of current knowledge, but also the likelihood of later retrieving that knowledge. With reference to "the largest desert on earth," an FOK judgment for the name "Sahara" would typically be assumed to involve a subjective feeling of knowing the answer and an estimate of the likelihood of recognizing or recalling it in the near future. However, as mentioned earlier, people might instead base their FOKs on the likelihood that they had once known the answer, whether or not they currently knew it. In this study, we asked whether the awareness of having once known an answer to a general-knowledge question reflects the same metacognitive process that contributes to judging the likelihood of later being able to recognize a response. If so, asking participants if they once knew the answer to a general-knowledge question should be analogous to traditional FOK measures. We have presented preliminary evidence that this is, in fact, the case.

We discussed toward the outset that FOK judgments reflect item familiarity, the retrievability or accessibility of the answer, and other metamemorial knowledge about the test item. It is noteworthy that Koriat and Levy-Sadot (2001) argued that the impact of the retrievability of full or partial information about an answer is regulated by item familiarity. In particular, they presented evidence that retrievability exerts its influence predominantly for high-familiarity items. The present study focused on the impact of familiarity on the duration of memory search, but it does not deny the possibility that retrievability might likewise influence these processes.

Taken together, these experiments clarify our understanding of people's decisions to either continue or terminate memory search about general-knowledge questions. As discussed at the beginning of the article, this relationship is not a straightforward one; that is, high FOK is alternately associated with fast responses (Koriat, 1993) or slow ones (Costermans et al., 1992; Nelson et al., 1984). These differences may reflect the method of eliciting FOK judgments. First, when FOKs are elicited only for unrecalled items (e.g., Nelson & Narens, 1980), higher FOKs are usually associated with longer RTs. This is likely because the conviction that one knows the answer motivates the continued search of memory. In contrast, when FOK evaluation is noncon-

tingently elicited for all items, higher FOKs are associated with shorter RTs. The present experiments augment these previous findings by indicating that the duration of memory search also depends on whether one decides the answer is known or not.

REFERENCES

COSTERMANS, J., LORIES, G., & ANSAY, C. (1992). Confidence level and feeling of knowing in question answering: The weight of inferential processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 142-150.

FILLMORE, C. J. (1968). The case for case. In E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory* (pp. 1-88). New York: Holt, Rinehart, and Winston.

FRIEDMAN, A., BROWN, N. R., & MCGAFFEY, A. P. (2002). A basis for bias in geographical judgments. *Psychonomic Bulletin & Review*, **9**, 151-159.

GLUCKSBERG, S., & MCCLOSKEY, M. (1981). Decisions about ignorance: Knowing that you don't know. *Journal of Experimental Psychology: Human Learning & Memory*, **7**, 311-325.

HART, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, **56**, 208-216.

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.

KENDALL, M. G., & STUART, A. (1967). *The advanced theory of statistics* (Vol. 2). Oxford: Hafner.

KING, J. F., ZECHMEISTER, E. B., & SHAUGHNESSY, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *American Journal of Psychology*, **93**, 329-343.

KLIN, C. M., GUZMÁN, A. E., & LEVINE, W. H. (1997). Knowing that you don't know: Metamemory and discourse processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 1378-1393.

KOLERS, P. A., & PALEF, S. R. (1976). Knowing not. *Memory & Cognition*, **4**, 553-558.

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., & 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.

LINDSAY, P. H., & NORMAN, D. A. (1977). *Human information processing: An introduction to psychology* (2nd ed.). New York: Academic Press.

METCALFE, J. (1986). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 288-294.

NELSON, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, **95**, 109-133.

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., LEONESIO, R. J., LANDWEHR, R. S., & NARENS, L. (1986). A comparison of three predictors of an individual's memory perfor-

mance: The individual's feeling of knowing versus the normative feeling of knowing versus base-rate item difficulty. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 279-287.

NELSON, T. O., & NARENS, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning & Verbal Behavior*, **19**, 338-368.

NELSON, T. O., & NARENS, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 26, pp. 125-140). San Diego: Academic Press.

RATCLIFF, R., & McKOON, G. (1989). Similarity information versus relational information: Differences in the time course of retrieval. *Cognitive Psychology*, **21**, 139-155.

REDER, L. M. (1982). Plausibility judgments versus fact retrieval: Alternative strategies for sentence verification. *Psychological Review*, **89**, 250-280.

REDER, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology*, **19**, 90-138.

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.

REDER, L. M., & SCHUNN, C. D. (1996). Metacognition does not imply awareness: Strategy choice is governed by implicit learning and memory. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 45-78). Mahwah, NJ: Erlbaum.

SCHREIBER, T. A. (1998). Effects of target size on feelings of knowing and cued recall: Implications for the cue effectiveness and partial-retrieval hypotheses. *Memory & Cognition*, **26**, 553-571.

SCOBORIA, A. R. (2004). Immediate and persisting effects of misleading questions and hypnosis on memory reports: An extension and replication. (Doctoral dissertation, University of Connecticut, 2005). *Dissertation Abstracts International*, **65**, 6673.

SINGER, M. (1981). Verifying the assertions and implications of language. *Journal of Verbal Learning & Verbal Behavior*, **20**, 46-60.

SINGER, M. (1984). Toward a model of question answering: Yes–no questions. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 285-297.

SINGER, M. (1986). Answering wh- questions about sentences and text. *Journal of Memory & Language*, **25**, 238-254.

SINGER, M. (1990). Answering questions about discourse. *Discourse Processes*, **13**, 261-277.

YOUNG, C. J. (2004). Contributions of metaknowledge to retrieval of natural categories in semantic memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 909-916.

NOTE

1. *Carmen* was composed by Georges Bizet. After the opera was critically condemned, Bizet became depressed and subsequently died. It was only afterward that the magnificence of *Carmen* was acknowledged.