Journal of Experimental Psychology: General
1995, Vol. 124, No. 3, 311-333

0096-3445/95/S3.00

Dissociating Knowing and the Feeling of Knowing: Further Evidence for the Accessibility Model

Asher Koriat
University of Haifa

Predictions derived from the accessibility model of the feeling of knowing (FOK; A. Koriat, 1993) were tested regarding the basis of FOK and the reason for its accuracy. According to the model, FOK monitors the accessibility of partial information about unrecallable targets, and its validity depends on the accuracy of that information. General knowledge questions were classified in terms of their tendency to precipitate answers in recall (accessibility, or ACC), and the proportion of such answers that were correct (output-bound accuracy, or OBA). FOK increased with increasing ACC independent of actual recognition memory, and the FOK-recognition correlation varied dramatically with OBA: It was positive for high-OBA questions, but nil or negative for low-OBA questions. The results suggest that people have no privileged access to the contents of their memory over and above what they can retrieve from it.

In this article I extend investigation of the accessibility model of the feeling of knowing (FOK) phenomenon proposed by Koriat (1993) to a situation calling for the retrieval of real-world knowledge from long-term memory. In the previous study, the evidence in support of this account rested primarily on a task tapping the short-term retention of nonsense strings. This somewhat artificial task has the advantage of affording a simple measure of the amount of correct and incorrect partial information accessible to a participant in terms of the number and correctness of the letters reported while minimizing the possible contribution of preexperimental variables. Here I test predictions of the accessibility model using a more typical task: retrieving information from long-term memory. The focus is on two questions: What is the basis for FOK? and What is the reason for its accuracy in predicting future memory performance?

The FOK phenomenon is best illustrated by the many everyday situations in which people try to recall the name of a person but fail to find it. These situations are sometimes accompanied by the subjective conviction that one knows the name and that one is likely to recall it given sufficient time and effort (see R. Brown & McNeill, 1966; Smith, 1994). Such memory blocking states have attracted much attention because they imply that participants can monitor the presence of information in the memory store that they are unable to retrieve (A. S. Brown, 1991; Nelson & Narens, 1990).

Much of the experimental work on FOK has concentrated on demonstrating its validity, that is, on showing that FOK judgments regarding a momentarily inaccessible target are diagnostic of its availability in the memory store. The impetus for this line of research came from Hart's pioneering studies in which FOK ratings after recall failure were found to predict the success of recognizing the correct target among distractors (Hart, 1965, 1967a, 1967b). This correlation between subjective and objective indexes of knowledge has been replicated in many subsequent studies (e.g., see Schwartz & Metcalfe, 1994). In most of these studies, a recognition memory test has been used as the criterion, but other criteria have also been explored, such as recall, perceptual identification, and relearning (e.g., Gruneberg & Monks, 1974; Nelson, Gerler, & Narens, 1984; Shimizu & Kawaguchi, 1993; for reviews, see Metcalfe & Shimamura, 1994; Schwartz, 1994).

There has been a greater concern in the past few years in specifying the basis of FOK. Research on this issue has been impeded, perhaps, by the commonly held view of FOK as being a "storage state indicator" (Hart, 1967a, p. 689). It was Hart who has explicitly put forward the *trace-access* view of FOK, and this view has since been implicitly endorsed in many discussions of FOK and tip-of-the-tongue (TOT) states (e.g., Yaniv & Meyer, 1987). According to the trace-access view, FOK represents the output of a specialized monitoring mechanism that can directly detect the presence of the target's trace in memory. Whenever a memory target is solicited, the internal monitor is consulted to determine whether the target is available in the memory store. If the monitor signals the presence of the target, then retrieval will be attempted. If, however, the monitor returns a negative value, a "don't know" report is issued, thus sparing the time and effort involved in searching for some-

A brief version of this article was presented at the annual meeting of the Psychonomic Society, St. Louis, Missouri, November 1994.

This research was supported by a grant from the Israel Science Foundation administered by the Israel Academy of Sciences and Humanities. The experiments were conducted at the Institute of Information Processing and Decision Making, University of Haifa. I want to thank Morris Goldsmith for his comments on an earlier version and Michal Sion, Galit Aharonovich, and Ravit Levy for their help in all stages of the research program.

Correspondence concerning this article should be addressed to Asher Koriat, Department of Psychology, University of Haifa, Haifa, Israel. Electronic mail may be sent via Internet to rsps301@uvm.haifa.ac.il.

thing that is not in store. As noted by Koriat (1994), the two-stage monitor-and-retrieve conception of memory search implicit in this account is analogous to the manner in which files are searched in computerized systems: The directory is scanned to ascertain that it contains the name of a file before an attempt is made to fetch the file itself.

One important aspect of the trace-access model is that it also addresses a second question about FOK: why FOK judgments are accurate in predicting subsequent recall or recognition performance. Clearly, if FOK directly monitors the presence of the target's trace in memory, then it ought to serve as a valid predictor of actual memory performance. In fact, the accuracy of FOK in predicting memory performance has been implicitly seen to constitute, in itself, supportive evidence for the trace-access view of FOK. Indeed, if this view is endorsed, the inaccuracy rather than the accuracy of FOK needs to be explained.

The more recent interest in the basis of FOK, however, has led to a consideration of other mechanisms underlying the phenomenon that do not presuppose direct access to the memory trace. These mechanisms share the assumption that FOK is based on an inferential process in which certain cues are used consciously or unconsciously to form an FOK judgment about the likelihood that the inaccessible target is "there" and will be recalled or recognized at some later time. Nelson et al. (1984) have provided a comprehensive review of possible inference-based mechanisms for FOK (see also Krinsky & Nelson, 1985; Miner & Reder, 1994; Schwartz, 1994). For example, FOK judgments may be based on the familiarity with the general topic in question (e.g., "I know very little about South Africa"; see Costermans, Lories, & Ansay, 1992; Glenberg & Epstein, 1987) or on the retrieval of pertinent episodic information (e.g., remembering the episode in which one first met the person whose name one fails to retrieve).

Among the inference-based accounts of FOK, the one that has attracted the largest amount of experimental work so far is the cue-familiarity account (e.g., Metcalfe, 1994; Metcalfe, Schwartz, & Joaquim, 1993; Miner & Reder, 1994; Reder, 1987; Reder & Ritter, 1992; Schwartz & Metcalfe, 1992). According to this account, FOK is determined by the familiarity of the cue presented to the participant. Following Koriat and Lieblich (1977), the term *memory pointer* is used here to designate any cue (e.g., a question or a stimulus word) that is intended to specify a particular memory entry. Thus, according to the cue-familiarity hypothesis, FOK judgments monitor the familiarity of the memory pointer rather than the availability or retrievability of the solicited target itself. Indeed, several results have been reported indicating that FOK judgments can be enhanced by advance priming of the cue but not by priming of the target (e.g., Metcalfe et al., 1993; Reder, 1987, 1988; Reder & Ritter, 1992; Schwartz & Metcalfe, 1992). In fact, cue priming has been found to enhance FOK judgments independent of actual memory performance.

The present study explored another inference-based account of FOK—the accessibility account—with the aim of demonstrating how FOK can be dissociated from actual knowing. According to the accessibility model proposed by Koriat (1993,1994), the cues for FOK reside in the products of the retrieval process itself. Whenever memory is searched for a solicited target, a variety of clues often come to mind (see Durso & Shore, 1991; Gardiner, Craik, & Bleasdale, 1973; Lovelace, 1987; Read & Bruce, 1982), including fragments of the target, semantic attributes, episodic information pertaining to the target, and activations emanating from other sources. Although such clues may not be articulate enough to support an analytic, calculated inference, they can still act en masse to produce the subjective feeling that the target is there and will be recalled or recognized in the future. FOK, then, is generally based on a nonanalytic inference (see Jacoby & Brooks, 1984) that considers the overall accessibility of partial information pertaining to the target. Essentially, FOK is assumed to rely on an attempt to extrapolate from the processes that occur during the early stages of one retrieval episode to future retrieval episodes: If a memory pointer activates many associations, it is likely to eventually lead to the recollection of the target, but if it leaves one "blank," chances are that it will continue to bring nothing to mind. This account of FOK resembles the availability heuristic postulated by Tversky and Kahneman (1973) to explain how people estimate proportions or frequencies.

The accessibility account of FOK contrasts with the trace-access model in two respects. First, unlike the trace-access model, which implies a modular organization of monitoring and retrieval (with monitoring preceding retrieval), the accessibility account assumes an interactive process that combines retrieval and monitoring: It is by attempting to search for the solicited target that one can judge the likelihood that the target resides in memory and is worth continuing to search for. FOK judgments, then, are computed and updated on line on the basis of clues accumulated during the initial stages of search and retrieval. These judgments can then feed back into the retrieval attempt by motivating or discouraging further search for the target (e.g., Barnes, Nelson, Dunlosky, Mazzoni, & Narens, 1994; Gruneberg, Monks, & Sykes, 1977; Nelson et al., 1984; Nelson & Narens, 1980). Thus, in terms of the distinction drawn by Tulving and Pearlstone (1966), FOK monitors accessibility rather than availability.

Second, because the monitoring process is not independent of the retrieval process, the accuracy of FOK in predicting memory performance is not guaranteed but should vary depending on the quality of the partial clues that come to mind. If the retrieval attempt goes astray, so will monitoring. For example, the retrieval process may often be fooled by misleading clues. These clues may originate from a wrong referent (Koriat, 1994; Nelson & Narens, 1990), from memory entries that are stored in the vicinity of the solicited target (Koriat & Lieblich, 1977), from "interlopers" that come to mind (see Jones, 1989; Reason & Lucas, 1984), from misleading postevent information (e.g., Weingardt, Leonesio, & Loftus, 1994), or from activations resulting from advance priming (e.g., Jacoby & Kelley, 1991; Kelley & Lindsay, 1993). Participants often cannot specify the source of these activations. Therefore, they cannot simply discount them by attributing them to their proper source

(Jacoby & Whitehouse, 1989; Jacoby, Woloshyn, & Kelley, 1989). Such clues are likely to lead to the illusion of knowing. Because FOK judgments do not have privileged access to any information that is not already contained in the output of the retrieval attempt, their accuracy should depend on the correctness of the partial clues that come to mind.

What are the implications of this account regarding the basis of FOK and its accuracy? With regard to the basis of FOK, it is predicted that FOK should increase with the overall accessibility of information, regardless of its source. Thus, both correct and incorrect clues should contribute equally to FOK. This assumption distinguishes the accessibility account from the *target retrievability* account of FOK (see Nelson et al., 1984; Schacter & Worling, 1985; Schwartz & Metcalfe, 1992), according to which FOK is based on partial recall of the target proper. The assumption underlying this latter account is that although participants sometimes fail to retrieve the entire (correct) target, they may retrieve parts of it, and these parts are sufficient to activate a positive FOK. This is why FOK judgments tend to be accurate in predicting subsequent recall or recognition of the target. The accessibility model, in contrast, assumes that participants cannot monitor directly the accuracy of the information that comes to mind during a retrieval attempt (although they may be able to infer it from the intensity of that information, such as its ease of access, vividness, or persistent recurrence; see Koriat, 1993). Therefore, both correct and incorrect partial clues should generally contribute to the enhancement of FOK.

With regard to the accuracy of FOK, the question that emerges is why FOK judgments are nevertheless accurate in monitoring memory performance if they are based on the mere accessibility of information. According to the accessibility account, the accuracy of metamemory is indeed a by-product of the accuracy of memory itself. But memory itself is generally accurate in the sense that the partial clues that come to mind have a greater likelihood of being correct than being incorrect. This is because, by and large, information that has been memorized is more likely to give rise to correct than to incorrect full or partial recalls (see Koriat, 1993). Therefore, a monitoring mechanism that relies on the mere accessibility of information is bound to be predictive of subsequent recall or recognition performance.

However, the quality (i.e., correctness) of the information that comes to mind may still differ from situation to situation and from one set of memory pointers to another. Such differences should be expected to affect the accuracy of FOK in predicting memory performance. For example, if the items of a memory test precipitate mostly correct partial information (as is generally the case), then FOK judgments should be generally accurate, because the test items will tend to differ primarily in the amount of correct information that they precipitate. On the other hand, if the test is contrived so as to include mostly "deceptive items" (i.e., items that produce more incorrect than correct answers; see Fischhoff, Slovic, & Lichtenstein, 1977; Koriat, 1976; Koriat & Lieblich, 1977; Nelson et al., 1984), then FOK judgments may prove faulty in predicting subsequent memory performance. In the extreme, when the amount of incorrect clues exceeds the amount of correct clues, FOK judgments not only will be markedly inflated relative to actual memory performance but may, in fact, correlate negatively with successful recall or recognition: The more one perceives that one knows, the less likely that one would actually recall or recognize the correct target.

Note that, according to this formulation, what matters for the accuracy of FOK is the *output-bound accuracy* (OBA) of the responses rather than their input-bound accuracy, to use the terminology of Koriat and Goldsmith (1994, in press). OBA is defined as the likelihood that an answer (or a partial clue) that comes to mind is correct, whereas input-bound accuracy refers to the difficulty of the item (i.e., the likelihood that it will bring to mind the correct answer). Assuming that FOK judgments increase as a function of the amount of information that comes to mind, the critical determinant of FOK accuracy should be the conditional probability that an answer that has come to mind is correct rather than the unconditional (input-bound) probability that a correct answer will come to mind.

The experiments reported here followed the item-based logic underlying the work reported by Koriat and Lieblich (1977). This logic was motivated by the somewhat surprising observation that memory pointers (word definitions) differ reliably in the extent to which they tend to evoke a TOT state across participants. This observation suggested that some insight into the nature of the FOK and TOT states can be gained by focusing on systematic differences between memory pointers. Indeed, the results of Koriat and Lieblich indicated that memory pointers do exhibit reliable differences across participants that can be described in terms of two orthogonal factors: effectiveness in suggesting or eliciting the correct target (objective knowledge) and degree of FOK (subjective knowledge).

The experiments reported, then, were based on the assumption that memory pointers differ reliably in terms of the amount and the quality of information they tend to precipitate. Furthermore, it was assumed that normative data about interpointer differences provide information about the processes that occur within each individual participant. As in Koriat and Lieblich's (1977) study, these experiments used questions for which the answer was a single word, either the name of a person or place or a specific term. For Experiment 1, a large set of such questions was compiled, questions likely to differ in terms of the two properties that are pertinent to FOK: the amount of information that they bring to mind (accessibility, or ACC) and the likelihood that this information is correct (OBA). An attempt was made to include a relatively large proportion of deceptive pointers (i.e., those likely to precipitate incorrect answers). Each memory pointer was then scored on both ACC and OBA, which were defined operationally in terms of the free-recall responses of the participants. ACC was defined as the percentage of participants who provided an answer to the question (regardless of whether the answer was correct or wrong), and OBA was defined as the percentage of correct responses among the responses provided (OBA can be computed only when ACC is greater than zero). These two indexes were then used to test predictions about the bases of FOK and its accuracy. FOK is expected to increase as a function of ACC, regardless of actual memory performance, and the accuracy of FOK is expected to increase with OBA.

In some of the tests of these predictions reported later, the analyses were confined to FOK judgments provided either after recall failure (Experiment 1) or before deliberate retrieval was attempted (Experiment 2). These tests were predicted on the assumption that the overt recall responses associated with a particular memory pointer are diagnostic of the covert responses activated by that pointer when retrieval of the complete target is thwarted. Thus, high-ACC items are assumed to elicit a large number of partial clues even when no overt response is provided. Similarly, OBA, which is based on the accuracy of the answers provided, is assumed to be diagnostic of the quality (correctness) of the partial clues that are accessed before a deliberate search is attempted or those left behind when retrieval of the answer fails.

The rationale for Experiment 1 bore some similarity to that of an experiment reported by Koriat (1976; see also Koriat, 1975), although that experiment concerned retrospective confidence judgments rather than prospective FOK judgments. In that study, English-speaking participants were asked to choose the English translation of words from noncognate languages and to indicate their subjective confidence in their choices. Although confidence judgments were positively correlated with the correct translations for a subset of the items for which most participants endorsed the correct translation, they were negatively correlated for a subset of the items in which participants' responses were predominantly incorrect. Thus, monitoring accuracy depends on the overall accuracy of the responses.

Experiment 1

For Experiment 1, 95 questions were compiled that were judged to vary widely in their overall likelihood of eliciting correct and incorrect answers. All questions required a one-word answer. Participants attempted to recall the answer to each question and then provided an FOK judgment about the likelihood of selecting the correct answer from among four alternatives. A recognition test was then administered. It was expected that FOK judgments would vary with the accessibility of information regarding the target in question and that the accuracy of these judgments in predicting recognition memory would depend on the accuracy of the answers that came to mind.

Method

Participants. Thirty-six psychology undergraduates participated in the experiment for course credit.

Stimulus materials. A 95-item general knowledge test was developed, with questions covering a broad range of topics. All questions required a one-word answer, either a concept or a name of a person or place. A deliberate attempt was made to include questions that differed widely in their likelihood of evoking correct and incorrect answers. The questions were compiled from many sources, taking advantage of some of the deceptive items reported in the literature (e.g., Fischhoff et al., 1977; Gruneberg, Smith, & Winfrow, 1973; Nelson et al., 1984). In addition, several pretests were carried out to ensure a sufficiently large number of deceptive items.

Two booklets were prepared, a recall version and a recognition version. They included the same questions, but whereas in the recall version two blank lines were provided next to each question (for recording the answer and the FOK judgment), in the recognition version each question was followed by four possible answers, of which only one was correct. The order of the questions differed in the two booklets.

Procedure. The experiment was administered in group sessions lasting about 40 min. All experimental materials and instructions (in Hebrew) were compiled into self-contained booklets for individual students, who were told to read the instructions and proceed at their own pace. The instructions stated that the students would be required to answer a series of general knowledge questions twice, once in an open-ended format and once in a four-alternative forced-choice format. For each question of the recall phase, the students were told to "write down the answer if you know it . . . and then assess the chances that you will be able to identify the answer among four distractors." Students indicated their judgments on a 25%-100% scale (the instructions explained that 25% constitutes chance performance). Unlike the common practice of eliciting FOK judgments only after recall failures (see Koriat, 1993), here FOK judgments were always solicited. After the recall booklets had been collected, students were given the recognition booklets and asked to circle one answer for each question.

Results

The first set of analyses to be reported concerns pointer properties. Mean percentages of correct recall ranged from 0.0 to 97.2 across the 95 memory pointers, and mean correct recognition percentages ranged from 8.3 to 97.2. Recall and recognition percentages averaged 26.2 and 54.6, respectively, across pointers.

Two scores were computed for each pointer on the basis of the percentage of correct recalls, commission errors, and omission errors precipitated by the pointer. The first, an ACC score, was defined as the percentage of students reporting an answer (whether right or wrong). ACC ranged from 5.6% to 97.2% across pointers and averaged 45.8%. Second, an OBA (see Koriat & Goldsmith, 1994) score was defined as the percentage of correct responses of all of the responses produced to that pointer (i.e., correct responses plus commission errors). This score ranged from 0.0 to 100.0. For example, the OBA score for the question "Who was the first Roman emperor?" was 0.0: Of the 14 participants who answered this question, none gave the correct answer (Augustus). In contrast, the question "What is the original family name of the singer Bob Dylan?" yielded an OBA score of 100.0: Only 6 participants answered this question, but all of them gave the correct answer (Zimmerman). The OBA score averaged 53.6 across all pointers. Thus, roughly speaking, the pointers used were, on average, equally likely to culminate in an omission or a commission response, and the answers reported were equally likely to be right or wrong. Examples of the pointers used and their ACC and OBA scores can be found in Table 1 (translated from Hebrew).

The ACC and OBA scores associated with each pointer are used in the following analyses to examine both the basis of FOK judgments and the basis of their predictive validity.

Basis of FOK judgments. According to the accessibility model, FOK judgments are based on the overall amount of information accessed about the target irrespective of the accuracy of that information. A crude index of accessibility is whether the search process culminated in some answer (correct or wrong) or in an omission error. For each student, mean FOK judgments for correct answers, commission errors, and omission errors were calculated, and these percentages averaged 94.8, 84.2, and 59.1, respectively, across students. Thus, FOK judgments were significantly higher when an answer was provided (90.5%) than when the trial culminated in an omission error, $F(1, 35) = 335.59$, $p < .0001$, and this was true both when the retrieved answer was correct, $F(1, 35) = 387.89$, $p < .0001$, and when it was wrong, $F(1, 35) = 186.73$, $p < .0001$. These results replicate those previously reported (see Krinsky & Nelson, 1985; Nelson & Narens, 1990). Note, however, that correct answers were nevertheless associated with higher FOK judgments than wrong answers, $F(1, 35) = 65.68$, $p < .0001$. The reasons for this finding are discussed later.

A finer index of accessibility is the ACC score. A comparison of the high-ACC and low-ACC pointers in terms of their mean FOK judgments has the advantage that it can be performed while the overt response (omission or commission) is held constant. The assumption is that high-ACC pointers engender a higher degree of activation overall than low-ACC pointers, regardless of the success of retrieving a complete answer.

Table 1
Examples of the Pointers Used and Their ACC and OBA Scores

Pointer class	Pointer	Answer	ACC score (%)	OBA score (%)
CC	What actress played Dorothy in the original version of the movie "The Wizard of Oz"?	Judy Garland	33.3	75.0
CC	What is the name of India's "holy" river?	Ganges	69.4	92.0
CC	What is the name of the jazz player known as "Bird"?	Charlie Parker	16.7	83.3
CC	What Austrian researcher discovered the laws of genetics?	Mendel	33.3	83.3
CC	In Greek mythology, who was punished by having to pull a stone to the top of a mountain over and over?	Sisyphus	58.3	95.2
CC	What word refers to the ability to move objects by thought?	Telekinesis	52.8	68.4
CW	What is the capital of California?	Sacramento	47.2	23.5
CW	Who is the famous European artist who painted "The Potato Eaters"?	Van Gogh	11.1	25.0
CW	What is the capital of Uganda?	Kampala	13.9	0.0
CW	In which U.S. state is Yale University located?	Connecticut	11.1	0.0
CW	Who composed "The Unfinished Symphony"?	Schubert	94.4	8.8
CW	Corsica island belongs to what country?	France	61.1	9.1
low-ACC	What is the largest of the 23 states of Brazil?	Amazonas	0.0	
low-ACC	Who invented peanut butter?	George Washington Carver	0.0	
low-ACC	Who won the Pulitzer Prize for photography in 1970?	Steve Star	0.0	
low-ACC	What is the capital of Fiji?	Suva	0.0	
low-ACC	What is Iceland's legislature called?	Alting	0.0	
low-ACC	What was Montreal's nickname in the years 1920–1940?	Sin City	0.0	

Note. The low-ACC pointers were used only in Experiment 3, and their accessibility scores are based on the results of Experiment 2. OBA = output-bound accuracy; CC = consensually correct; CW = consensually wrong; low-ACC = low accessibility.

For the following analyses, all pointers were divided at the median ACC score between those yielding 15 (ACC = 41.7) or fewer answers (49 pointers) and those yielding 16 (ACC = 44.4) or more answers (46 pointers) overall. Mean FOK judgments associated with each of the two classes of pointers were calculated separately for commission and omission trials (see Figure 1). High-ACC pointers engendered higher FOK judgments than low-ACC pointers for both commission trials, $F(1, 35) = 18.61$, $p < .0001$, and omission trials, $F(1, 35) = 62.19$, $p < .0001$. Thus, even when the overt response (commission or omission) is held constant, the FOK judgments associated with a given pointer can be reliably predicted from the percentage of participants for whom that pointer precipitated some answer.

The interpretation of these results is complicated by the fact that the percentage of correct responses was higher for the high-ACC (59.9) than for the low-ACC (47.7) pointers, and, as reported earlier, correct responses were associated with higher FOK judgments than incorrect responses. Therefore, two sets of 37 pointers each were selected that differed in overall accessibility but were matched on OBA. ACC averaged 25.4% and 64.5% for the low- and high-ACC sets, respectively, whereas OBA averaged 54.4% and 53.8%, respectively. Figure 1 also presents the results of analyses based only on these 74 pointers. Comparisons of only commission trials revealed that high-ACC pointers yielded a higher FOK than low-ACC pointers, $F(1, 35) = 4.44$, $p < .05$. Note that the corresponding recognition performance was no better for high-ACC (61.3%) than for low-ACC (67.1%) pointers, $F(1, 35) = 3.16$, *ns*. Similarly, in comparisons of only omission trials, high-ACC pointers yielded higher FOK judgments than low-ACC pointers, $F(1, 35) = 42.38$, $p < .0001$. The respective means for recognition performance were 50.1% and 44.2%, $F(1, 35) = 3.49$, *ns*. Thus, when OBA is controlled, accessibility seems to specifically affect FOK without affecting recognition memory.

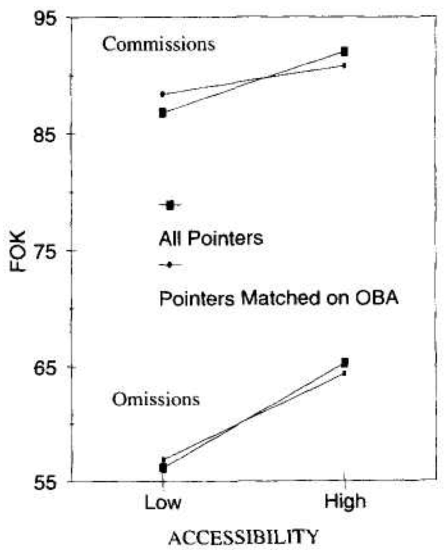

Figure 1. Mean feeling of knowing (FOK) judgments for low-accessibility and high-accessibility pointers in commission and omission trials. The results are presented separately for an analysis that included all pointers and for an analysis including only pointers matched on output-bound accuracy (OBA; Experiment 1).

The dissociation demonstrated here between FOK and recognition memory performance was obtained in a pointer-based analysis and agrees with the observation of Koriat and Lieblich (1977) that knowing and FOK emerge as two independent factors in an item-based analysis of memory pointers. A similar dissociation has been demonstrated in studies concerning the effects of cue familiarity on FOK (e.g., Metcalfe et al., 1993; Reder & Ritter, 1992; Schwartz & Metcalfe, 1992; see also Jameson, Narens, Goldfarb, & Nelson, 1990; Narens, Jameson, & Lee, 1994). I now turn to the within-individual correlations between subjective and objective indexes of knowing (i.e., to the predictive accuracy of FOK).

Accuracy of FOK in predicting recognition performance. The accuracy of FOK in predicting recognition memory performance was evaluated by dividing each student's FOK judgments at the median and comparing the recognition percentages for below-median and above-median FOK judgments for that student. (Six students had a median of 100; for these students, the division used was between 100 and less than 100.)

Overall, students were relatively accurate in predicting their recognition performance: Across all 95 pointers, recognition performance for low-FOK and high-FOK judgments averaged 43.1% and 68.3%, respectively, $F(1, 35) = 251.79$, $p < .0001$. The within-subject gamma correlation (see Nelson, 1984) averaged .49, which was significantly different from zero, $t(35) = 16.24$, $p < .0001$. All 36 students evidenced better recognition for high-FOK than for low-FOK pointers ($p < .0001$ by a binomial test). These results are generally consistent with previous findings. Note, however, that the respective FOK means for this comparison were 53.6% and 95.6%, indicating that students were markedly overconfident (see Lichtenstein, Fischhoff, & Phillips, 1982).

According to the accessibility model, the accuracy of FOK should be moderated by the overall accuracy of the responses that come to mind when searching for a target. Therefore, the analyses reported earlier were repeated, and pointers that engendered predominantly correct answers and those that precipitated predominantly incorrect answers were separated across students. For these analyses, all pointers were divided according to their OBA scores. For 6 pointers, OBA was exactly 50%, and these pointers were not included in the analyses reported here. The remaining pointers were divided between a consensually correct (CC) class, which included 52 pointers with an OBA higher than 50%, and a consensually wrong (CW) class, which included 37 pointers with an OBA lower than 50% (the latter are also referred to as *deceptive* henceforth). It should be pointed out that there was a large variation among the deceptive pointers in the number of different types of incorrect answers they elicited across students. For 3 deceptive pointers, there was a single incorrect answer that predominated across students (e.g., the question "What is the capital of Holland?"); other pointers precipitated as many as nine different incorrect answers (e.g., the question "In what state is Yale University located?"). Examples of the CC and CW pointers appear in Table 1.

FOK accuracy was assessed separately for the CC and CW pointers with omission and commission trials combined. Figure 2 presents mean recognition performance for low-FOK and high-FOK responses for the CC and CW pointers. The results are plotted as a function of the actual FOK means. Also depicted in Figure 2 is the recognition performance that would be expected when calibration is perfect (see Lichtenstein et al., 1982). It can be clearly seen that the FOK—recognition correlation reported earlier for the total sample of pointers actually conceals a marked difference between the CC and CW classes. A two-way OBA Class (CC vs. CW) X FOK Level (below vs. above median) analysis of variance (ANOVA) on recognition memory yielded the following results: OBA class, $F(1, 35) = 423.21$, $p < .0001$; FOK level, $F(1, 35) = 104.47$, $p < .0001$; the OBA Class X FOK Level interaction, $F(1, 35) = 97.30$, $p < .0001$.

A separate analysis for the CC class yielded the following

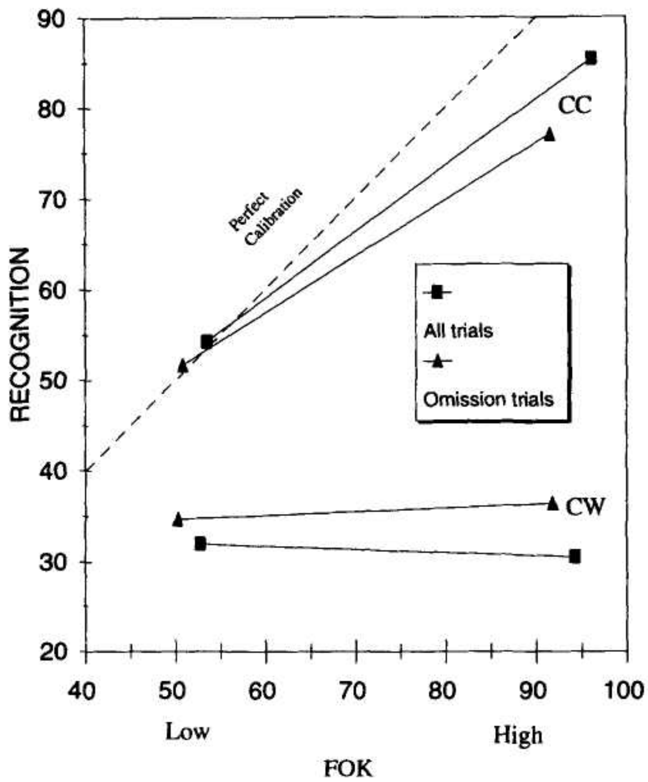

Figure 2. Mean recognition performance for low-FOK and high-FOK judgments for the consensually correct (CC) and consensually wrong (CW) pointers, plotted separately for both omission and commission trials combined and for omission trials only (Experiment 1). FOK = feeling of knowing.

result for FOK level: $F(1, 35) = 342.25$, $p < .0001$. FOK judgments for this class evidenced only a small degree of overconfidence, approaching perfect calibration (see Figure 2). All 36 students exhibited a trend indicating better recognition performance for high-FOK than for low-FOK pointers ($p < .0001$ by a binomial test). In contrast, for the CW class, FOK judgments evidenced a very marked discrepancy between subjective and objective knowledge: Whereas FOK judgments averaged 94.3% for the high-FOK pointers, recognition performance averaged 30.5%, barely better than chance. In addition, these judgments were totally unrelated to recognition performance ($F < 1$). In fact, 21 students' mean recognition performance was worse for high-FOK than for low-FOK pointers; 15 students exhibited the opposite trend (the difference was not significant by a binomial test). Thus, only for the CC class did recognition performance increase with increasing FOK judgments, whereas for the CW class the results, if anything, point in the opposite direction.

This conclusion is further substantiated by within-subject gamma correlations (see Nelson, 1984). The mean gamma correlation between dichotomized FOK judgments and recognition memory was .69 for the CC pointers, $t(35) = 23.98$, $p < .0001$ (for the difference from zero). For the CW pointers, in contrast, the correlation was -.05, $t(35) = 0.82$, *ns*.

The preceding results were based on all responses, regardless of whether students reported an answer or not. As has been argued before (Koriat, 1993), the common practice of eliminating pointers for which participants provided some answer (correct or incorrect) when estimating FOK accuracy is inappropriate from the point of view of the accessibility model. In the present context, however, the inclusion of commission responses in the analysis may create a methodological problem, because these responses contribute both to the classification of the pointers (as CC or CW) and to the evaluation of FOK accuracy. Statistical independence can be achieved, however, if FOK accuracy is evaluated with only pointers for which the participant failed to provide any answer (i.e., omission responses). Such an analysis, in a sense, uses the answers provided by those participants who supplied an answer to make predictions regarding the validity of FOK judgments made by those who failed to find an answer.

The analyses reported earlier were therefore repeated, but only omission trials were included. The results based on all 95 pointers combined indicated relatively good accuracy, with low-FOK and high-FOK pointers yielding 43.5% and 64.1% recognition performance, respectively, $F(1, 35) = 39.45$, $p < .0001$. The respective FOK means for this comparison were 51.1% and 91.9%, indicating, again, a high degree of overconfidence. The mean gamma correlation was .39, $f(35) = 6.47$, $p < .0001$. The correlation was positive for 29 students and negative for 7 students ($p < .0005$ by a binomial test).

A separate analysis was carried out for the CC and CW pointers, and the pertinent means also appear in Figure 2. (It should be noted that the CC means were based only on 35 students and the CW means were based only on 33 students for whom both high-FOK and low-FOK means were available.) The FOK means for the omission responses were lower than those obtained for the entire sample, as noted earlier, but they also yielded very different effects for the CC and CW pointers. An OBA Class X FOK Level ANOVA on recognition performance yielded the following results: OBA class, $F(l, 35) = 58.00$, $p < .0001$; FOK level, $F(l, 35) = 14.58$, $p < .0005$; and the OBA Class X FOK Level interaction $F(1, 31) = 32.33$, $p < .0001$.

A separate analysis for the CC class yielded the following result for FOK level: $F(1, 34) = 50.22$, $p < .0001$. Within-subject gamma correlations for this class averaged .50, $f(34) = 7.19$, $p < .0001$. The gamma correlation was positive for 29 students and negative for 5 students (it was zero for 1 student; $p < .0001$ by a binomial test).

In contrast, for the CW class, recognition performance was unrelated to FOK ($F < 1$). The mean gamma correlation for this class was negative, — .05, but not significantly different from zero, $r(32) = 0.41$. Gamma was positive for 17 students and negative for 16 students. Note again that, for the high-FOK responses, a mean FOK of 91.8% was associated with a 36.3% recognition performance.

Discussion

The results of Experiment 1 bear on the two questions addressed here (the basis of FOK and the basis of its predictive validity). As far as the basis of FOK is concerned, the results of Experiment 1 provide two lines of evidence in support of the proposition that FOK rests on the overall accessibility of information about the target, regardless of the accuracy of that information. First, FOK judgments were markedly higher after commission than after omission responses, and this was the case whether the answer produced was correct or wrong. This result replicates previous observations (Krinsky & Nelson, 1985; Nelson & Narens, 1990). It should be pointed out, however, that in previous studies FOK judgments have not been typically collected after correct responses because of the implicit assumption that there is no sense in soliciting FOK judgments when participants "know" the answer (see Koriat, 1993). However, the present results suggest that commission errors are, if anything, more similar to correct responses than to omission errors in terms of their implications for FOK judgments. Apparently, the mere accessibility of a pertinent answer is sufficiently potent subjective evidence that one knows the correct answer.

However, FOK judgments were nevertheless higher after correct commissions than after incorrect commissions, suggesting that the students may have had some access to the accuracy of their memory products. This finding can be taken to support the trace-access view of FOK. Alternatively, students may have inferred the accuracy of their memory products from additional cues involving intensity factors (e.g., the ease of accessing the answer). Indeed, several authors have proposed that ease of accessing information is a potent cue for metacognitive judgments such as judgment of learning (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989), FOK (Koriat, 1993), and subjective confidence (Kelley & Lindsay, 1993). Koriat (1993) found evidence suggesting that participants can monitor the accuracy of the partial or complete information retrieved. However, he also observed that correct information is retrieved with greater ease (i.e., shorter latency) than incorrect information and that FOK judgments are positively correlated with ease of access. These results suggest that ease of access may be the mediating factor underlying the relationship between FOK judgments and the accuracy of the retrieved information. Perhaps in the present study as well, correct answers were retrieved with greater ease than incorrect answers, and that is why they were also associated with higher FOK judgments.

The second line of evidence comes from the comparisons between low-ACC and high-ACC pointers: Pointers that precipitated a high rate of commission responses across students were associated with higher FOK judgments than pointers engendering few answers. This was true even when the comparison was confined to omission responses. Possibly even when retrieval of the target fails, high-ACC pointers activate more partial information than low-ACC pointers. Interestingly, these differences in covert accessibility were not correlated with actual recognition performance across the sample of pointers included in the experiment.

The second question addressed by the experiment concerns the mechanism underlying FOK's accuracy in predicting recognition performance. The results are generally consistent with the proposition that FOK accuracy is a byproduct of memory accuracy: Because FOK is based on the mere accessibility of information, it tends to be accurate as long as the information accessed is predominantly correct. This was indeed the case with the CC items, which yielded the commonly found positive correlation between FOK judgments and recognition memory. The correlation disappeared, however, in the case of the CW items, for which the information accessed was predominantly incorrect. Here not only were FOK judgments completely unpredictive of memory performance, but the assessed recognition probabilities were also overly inflated in comparison with the actual recognition probabilities. Apparently, the misleading activations emanating from the incorrect referents and their associates resulted in illusory feelings of knowing (see Koriat & Lieblich, 1977), much the same way that deceptive items engender an illusion of certainty in confidence studies (Fischhoff et al., 1977; Koriat, 1976).

In fact, a negative correlation between FOK and recognition performance was expected for the class of deceptive pointers (see Koriat, 1976). One possible mundane reason why this was not so is that, in the case of deceptive pointers, participants whose search process is guided by the correct referent are those who may also have access to the kind of specific information that contributes to the enhancement of FOK judgments (e.g., a participant who has lived in Canberra). This implies that some of the high-FOK judgments obtained with deceptive items may have been based on the content of the partial clues retrieved rather than merely on their accessibility. According to Koriat (1993), content considerations tend to enter into the computation of FOK judgments relatively late in the retrieval process. If such is the case, perhaps the negative FOK-recognition correlation expected for deceptive items would be obtained in a speeded monitoring task like that used by Reder and her associates (Reder, 1987,1988; Reder & Ritter, 1992). Also, the overall recognition performance for deceptive items was not much below chance level, as would be expected (see Fischhoff et al., 1977), possibly because of the use of a four-alternative recognition test. Experiment 2, then, used a speeded monitoring task and a two-alternative forced-choice recognition test, which allowed a greater space for performance to deviate from chance level.

Experiment 2

Experiment 2 applied a procedure similar to the "game-show" technique of Reder and her associates (see Miner & Reder, 1994), in which participants are urged to provide a fast FOK judgment before attempting to search deliberately for the answer. This procedure allows investigation of the accessibility model to be extended to the type of preliminary FOK assumed to precede directed search for the target.

The notion of a preliminary FOK that precedes retrieval has been advanced by several researchers who view FOK as a rapid and automatic judgment that is based on a shallow analysis of the memory pointer (e.g., Koriat & Lieblich, 1977; Metcalfe, 1993, 1994; Miner & Reder, 1994). Metcalfe, for example, regarded the FOK as the output of a novelty monitoring mechanism that assesses rapidly the familiarity or novelty of the pointer without regard to its content. Reder and her associates (Miner & Reder, 1994; Reder, 1987, 1988; Reder & Ritter, 1992) argued for a conception of FOK as a general mechanism that operates at a preretrieval stage and guides the choice of question-answering strategy. Indeed, Reder (1987, 1988) observed that less time was needed for making FOK judgments about the recallability of an answer than for retrieving the answer itself. She proposed that FOK judgments monitor cue familiarity rather than target accessibility.

These ideas imply a distinction between monitoring and retrieval (Reder & Ritter, 1992) and, in fact, concur with the trace-access view of FOK in assuming a modular organization in which monitoring precedes retrieval (see Koriat, 1994). In the accessibility model, in contrast, monitoring and retrieval are seen to be intermingled from the beginning, with FOK judgments being based on the by-products of the retrieval attempt. This model, then, favors a more continuous view in which retrieval is seen to encompass both the early inspection of the automatic activations emanating from the terms of the pointer and the deliberate consideration of alternative candidates. In such a view, both the preliminary FOK elicited before directed search for the target and the FOK elicited after recall failure are seen to be based on the same type of cue: the accessibility of partial information.

If the accessibility view is correct, it may be expected that preliminary FOK judgments elicited from participants before they have had sufficient time to search deliberately for the target should yield the same pattern of results as that observed in Experiment 1 for students who failed to reach the answer. Thus, ACC and OBA, as indexed by the number and quality of answers provided by participants who are allowed sufficient time to retrieve an answer (Experiment 1), should predict preliminary FOK judgments and their accuracy in Experiment 2.

The procedure of Experiment 2 was similar to that of Experiment 1 except that participants were read the general information questions and, for each question, were asked to provide a fast FOK judgment within a 5-s time limit. In addition, two modifications were introduced. The first concerned the selection of pointers. The pointers used in Experiment 1 included more CC than CW pointers, and the two sets were not entirely matched on accessibility, resulting in a correlation of .22 ($p < .05$) between ACC and OBA across all items. In Experiment 2, three sets of pointers were included: CC, CW, and low-ACC. The CW pointers were the 37 deceptive pointers used in Experiment 1, whereas the CC pointers were selected from those of Experiment 1 so that they matched the CW pointers in terms of the ACC index. In addition, because the classification of pointers as CC or CW required that they elicit some answers in recall, the pointers included in these two sets can be considered to represent a moderate to high accessibility level. Therefore, a set of 36 low-ACC pointers was added, comprising pointers that tended to evoke very few free-recall responses overall. These low-ACC pointers were expected to yield lower FOK judgments than either the CC or CW pointers. In this manner, it was possible to achieve a wider range of accessibility values while also allowing different degrees of OBA to be represented within the moderate to high accessibility level. Note that whereas in Experiment 1 both the recall data (which provided the basis for the classification of pointers in terms of ACC and OBA) and the FOK data were secured from the same sample of students, Experiment 2 achieved a greater degree of independence by using the results of Experiment 1 to define the CC and CW classes and by confining the primary analyses to omission responses.

Second, a two-alternative rather than a four-alternative recognition test was used as the criterion test in Experiment 2. As noted earlier, a four-alternative test is less likely to reveal the below-chance performance characteristic of deceptive pointers (see Fischhoff et al., 1977). Thus, perhaps the two-alternative format would bring to the fore the negative correlation between FOK and recognition that is expected for the CW pointers.

Method

Participants. Thirty psychology undergraduates participated in the experiment for course credit.

Stimulus materials. The recall data from Experiment 1 provided the basis for the selection of the CW and CC pointers used in Experiment 2. All 37 CW pointers were used, and 37 CC pointers were chosen to match them in terms of overall accessibility. Mean OBA scores for the two sets were 18.5 and 78.1, respectively, whereas mean ACC scores were 39.7 and 41.4, respectively. In addition, a set of 36 pointers was compiled, all requiring a one-word answer; these pointers were found, on the basis of preliminary testing, to evoke few free-recall responses, whether correct or incorrect. The foil alternatives in the recognition test for the CC and CW pointers were chosen from the most frequent incorrect answers produced in the recall phase of Experiment 1. For the low-ACC pointers, a response alternative judged to be a misleading foil was used. Examples of the low-ACC pointers used appear in Table 1.

Procedure. The experiment was administered in groups of 2-6 students. The procedure was the same as that of Experiment 1 with the following exceptions. In the first phase, students were told that they would be read a series of questions and that, for each question, they should provide a fast, preliminary assessment of the probability (on a 50%-100% scale) that they would be able to choose the correct answer from among two alternatives. It was indicated that students should not make a deliberate effort to search for the answer but that, if the answer came to mind within the 5-s period allotted, they should write it down next to the probability estimate. Each question was then read aloud by the experimenter, and students wrote down their FOK judgment on the blank space next to the number of the question read. A buzzer sounded at the end of the 5-s answering period, and the next question was then presented.

Results

Students provided a total of 136 answers across the CC pointers, of which 89.7% were correct, and 80 answers across the CW pointers, of which only 25% were correct. For the low-ACC pointers, only 3 answers were provided, of which 1 was correct. These results are consistent with the classification of the pointers as CC, CW, and low-ACC. FOK judgments were significantly higher for commission trials (98.3%) than for omission trials (70.0%), $F(1, 16) = 775.49$, $p < .0001$ (the analysis was based only on 17 students who produced 1 or more answers). For commission trials, correct answers were associated with an average FOK of 99.1%, in comparison with 98.0% for incorrect answers, $F(1, 14) = 1.64$, *ns.* All commission trials (which constituted 6.6% of all trials) were eliminated from the analyses reported here.

Effects of accessibility on FOK. I first examine the effects of accessibility on FOK judgments. The initial analysis compared FOK judgments for the three classes of pointers (CC, CW, and low-ACC). The mean FOK judgments for these classes were 76.1%, 76.4%, and 58.7%, respectively. Thus, first, the low-ACC class engendered considerably lower FOK than the two relatively high-accessibility classes combined (76.2%), $F(1, 29) = 427.80$, $p < .0001$, consistent with the idea that preliminary FOK increases with the amount of information accessible. Second, the CC and CW classes yielded practically identical FOK means ($F < 1$), supporting the view that FOK judgments do not monitor directly the accuracy of the accessible information.

A second analysis was based only on the 74 pointers (CC and CW) for which ACC estimates could be secured from the results of Experiment 1. As a means of examining the effects of ACC, these pointers were divided into those yielding 13 answers (ACC = 36.1%) or fewer in Experiment 1 (19 CC and 19 CW pointers; moderate ACC) and those yielding 14 (ACC = 38.9%) or more (18 CC and 18 CW pointers; high ACC). Mean FOK judgments were calculated for each student for each of the four groups of pointers, and the averages of these means are depicted in Figure 3. A two-way ANOVA on these data yielded the following results: OBA class (CC vs. CW), $F < 1$; ACC, $F(1, 29) = 134.81$, $p < .0001$; and the OBA Class X ACC interaction, $F(1, 29) = 10.39$, $p < .01$. FOK judgments increased with increasing ACC for both the CW class, $F(1, 29) = 33.09$, $p < .0001$, and the CC class, $F(1, 29) = 162.15$, $p < .0001$. Note, however, that the effects of ACC were stronger for the CC than for the CW class. This difference may perhaps derive from systematic differences in the ease of access of partial information. Koriat (1993), for example, observed that a correct piece of partial information makes a stronger contribution to FOK than an incorrect piece.

In general, when the CC and CW classes were combined, FOK judgments for the low-, moderate-, and high-ACC pointers averaged 58.7%, 72.3%, and 80.7%, respectively, $F(2, 58) = 323.86$, $p < .0001$. In sum, the results are consistent with the idea that the accessibility of pertinent information, irrespective of its accuracy, is a powerful determinant of preliminary FOK.

Determinants of FOK accuracy. I now turn to the analyses pertaining to FOK accuracy. Consider first the comparisons among the three classes of pointers. Figure 4 depicts mean recognition performance for these classes as a function of their mean preliminary FOK judgments. The figure also indicates the recognition performance that would ensue from perfect calibration. The results demonstrate a double dissociation between objective and subjective in-

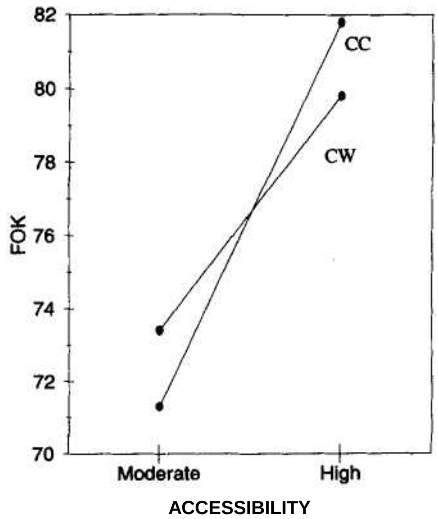

Figure 3. Mean preliminary feeling of knowing (FOK) judgments in Experiment 2 for moderate-accessibility and high-accessibility pointers, plotted separately for consensually correct (CC) and consensually wrong (CW) pointers (classified on the basis of the results of Experiment 1).

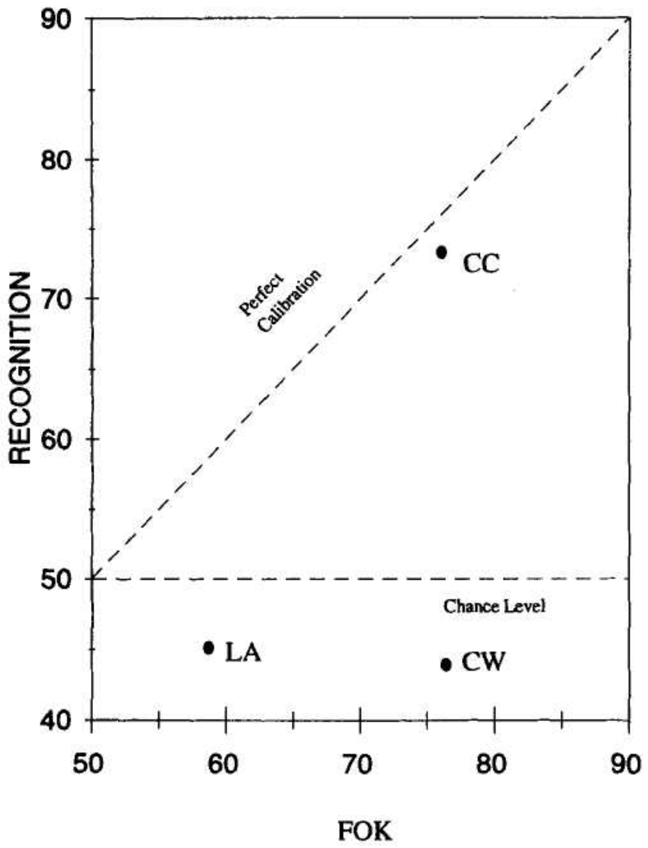

Figure 4. Overall mean recognition memory for three classes of pointers, consensually correct (CC), consensually wrong (CW), and low accessibility (LA), plotted as a function of their corresponding mean preliminary feeling of knowing (FOK) judgments (Experiment 2).

dexes of knowing: Although recognition memory was considerably better for the CC than for the CW pointers, the two classes of pointers evoked practically identical preliminary FOK judgments. On the other hand, the CW pointers were associated with considerably higher FOK judgments than the low-ACC pointers, even though both classes yielded similar recognition performance.

The equivalent recognition performance for the CW and low-ACC classes is somewhat surprising. The low-ACC pointers were expected to yield slightly above-chance recognition performance, whereas the CW pointers were expected to yield slightly below-chance performance. However, because of the procedure used in selecting the distractors for the low-ACC items, in many of the low-ACC pointers the wrong alternative was apparently more attractive than the correct alternative (e.g., for five pointers the incorrect distractor was chosen by 80% or more of the students), resulting in the low recognition performance for the low-ACC class.

I turn next to a within-subject analysis of the predictive validity of FOK. For this analysis, each student's FOK judgments were split at the median. Across all 110 pointers, mean recognition performances for low-FOK (below the median) and high-FOK (above the median) responses were 50.3%, and 58.8%, respectively, $F(l, 29) = 15.24$, $p < .001$. Thus, overall FOK judgments effectively predict recognition performance. It should be mentioned that the respective FOK means were 56.8% and 88.3%, again indicating considerable overconfidence (see also Figure 4). The mean gamma correlation (.15) between dichotomized FOK judgments and recognition was low but significant, $t(29) = 3.32$, $p < .005$ (for the difference from zero). Gamma was positive for 23 students, and negative for 7 students ($p < .005$ by a binomial test).

As in Experiment 1, however, these results conceal important differences between the different classes of pointers, as can be seen in Figure 5. The results in Figure 5 are plotted as a function of the actual FOK means (as in Figure 2). A two-way, Class (CC vs. CW vs. low-ACC) X FOK Level (below vs. above median) ANOVA yielded the following results: class, $F(2, 58) = 80.83$, $.p < .0001$; FOK level, $F(l, 29) = 1.19$, *ns;* and the Class X FOK Level interaction, $F(2, 58) = 14.36$, $p < .0001$. Separate one-way ANOVAs for each of the three classes of pointers indicated that the effects of FOK level were significant for both the CC class, $F(l, 29) = 37.13$, $p < .0001$, and the CW class, $F(l, 29) = 13.98$, $p < .0001$, but not for the low-ACC class ($F < 1$). However, whereas recognition performance for the CC pointers increased significantly with increasing FOK, recognition performance for the CW pointers decreased significantly as FOK increased. Note that the range of FOKs represented by the low-FOK and high-FOK pointers was quite wide for all three classes (although, of course, the low-FOK means for the low-ACC pointers were based on a larger number of observations than the high-FOK means).

These conclusions were substantiated by within-subject gamma correlations. The mean gamma correlation between dichotomized FOK judgments and recognition was .31 for the CC pointers, $f(29) = 4.30$, $p < .0005$ (for the difference from zero). For the CW pointers, in contrast, it was significantly negative (-.18), $t(29) = 3.13, p < .005$. The gamma correlation for the CC pointers was positive for 23 students and negative for 7 students ($p < .005$ by a binomial test), whereas the correlation for the CW pointers was positive for 6 students and negative for 24 students ($p < .001$). The gamma correlation for the low-ACC pointers averaged .02 across students, $f(29) = 0.18$, *ns.*

In sum, the FOK-recognition correlation is clearly contingent on the nature of the pointers included in the sample. Although across all pointers FOK judgments were generally accurate, the FOK-recognition correlation was positive for the CC pointers, negative for the CW pointers, and practically zero for the low-ACC pointers.

Some insight into the process underlying illusory preliminary FOK can be gained from a comparison of the deceptive pointers that precipitated many different answers across students and those that elicited only a few. Assuming that a cross-subjects variability is indicative of within-subject variability, the former type of pointers would be expected to evoke more conflicting clues than the latter. If FOK is based

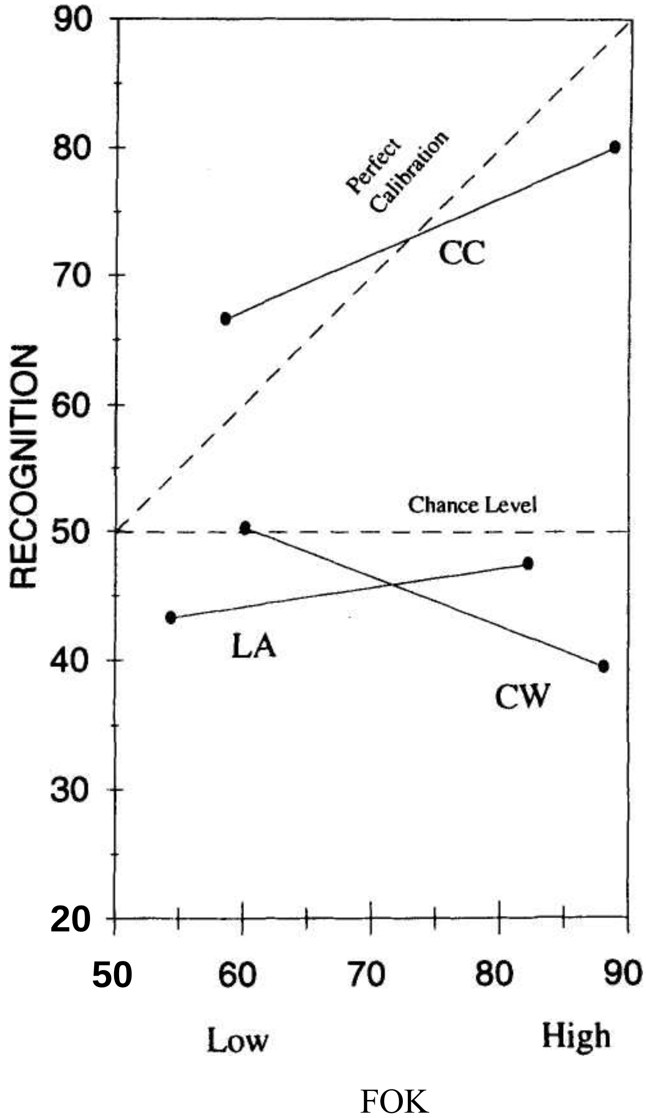

Figure 5. Mean recognition performance for preliminary low-FOK and high-FOK judgments, plotted separately for consensually correct (CC), consensually wrong (CW), and low-accessibility (LA) pointers (Experiment 2). FOK = feeling of knowing.

on an analytic weighing of the evidence, such pointers should, perhaps, be associated with lower FOK judgments. This, however, was not the case: When all deceptive items were classified into those eliciting up to three incorrect answers across all students in Experiment 1 (n = 18) and those eliciting four or more (n = 19), FOK judgments in Experiment 2 averaged 77.5% and 78.4% for these two groups, respectively, $F(1, 29) = 1.50$, *ns*. This lack of effect, then, suggests that FOK judgments elicited in the early stages of retrieval are determined by a nonanalytic process that considers the mere accessibility of clues regardless of their content (Koriat, 1993, 1994). Therefore, the compatibility between different clues (e.g., the possibility that they may be in conflict) may not be taken into account in the forming of a preliminary FOK judgment. These results are somewhat at odds with those reported by Schreiber and Nelson (1994), who found that cues linked to larger sets of associates produce lower FOK ratings than cues linked to smaller sets. However, their procedure differed from that of the present study, and more research is needed to determine whether the discrepancy between the two sets of findings is real.

Discussion

Experiment 2 extended investigation of the accessibility model to a situation tapping preliminary FOK judgments. The results were generally consistent with predictions, indicating that the classification of pointers in terms of their recall characteristics is predictive of both the strength and the accuracy of preliminary FOK judgments before a deliberate search for the target.

Consider first the effects of ACC. The CC and CW pointers were matched on ACC and, indeed, engendered equivalent preliminary FOK judgments (76.1% and 76.4%, respectively), even though they differed considerably in both recall performance (in Experiment 1; 32.4% and 7.2%, respectively) and recognition performance (in Experiment 2; 73.3% and 43.9%, respectively). These results clearly testify for the claim that FOK judgments do not monitor the accuracy of accessible information.

On the other hand, both classes were found to evoke considerably higher FOK judgments (76.2%) than the low-ACC pointers (58.7%), which presumably evoke few answers, suggesting a key role for accessibility as such. Note that recognition performance was very similar for the CW and low-ACC pointers (43.9% and 45.1%, respectively), although the former were found to yield considerably higher FOK judgments. In terms of FOK accuracy, the results clearly demonstrated that the correlation between FOK and recognition memory can be positive, negative, or zero depending on the recall characteristics of the set of items included in the sample.

Apart from providing further support for the accessibility model of FOK, the results of Experiment 2 also suggest that the preliminary FOK preceding deliberate search and the FOK that occurs when that search fails have a common basis: the amount of partial clues precipitated by the pointer. The finding that fast, preliminary FOK judgments can be predicted from the number of answers produced by participants who have had sufficient time to retrieve an answer argues against the notion of preliminary FOK as emanating from a preretrieval stage (e.g., Reder & Ritter, 1992). Rather, this finding favors a conception in which the initial, shallow analysis of the pointer is seen to represent an integral part of the process of answering a question. The fast, preliminary FOK would then be seen to monitor the overall amount of clues precipitated by the shallow analysis of the pointer in much the same way that the FOK elicited after a recall failure is seen to monitor the amount of scattered debris left behind by the abortive attempt to retrieve the target.

Furthermore, not only was the level of preliminary FOK predictable from the number of answers provided in Experiment 1, but the accuracy of preliminary FOK was correlated with the correctness of these answers. This would seem to argue against the notion that FOK judgments are based strictly on the mere familiarity of the terms of the pointer (e.g., Metcalfe, 1994; Miner & Reder, 1994).

Experiment 3

The results of Experiments 1 and 2 indicated that FOK judgments associated with a target that is not immediately recallable can be predicted from the number of answers provided by participants who did reach an answer. Furthermore, the accuracy of these FOK judgments can be predicted from the accuracy (correctness) of the answers produced by those participants who succeeded in reaching an answer.

The interpretation of these results was based on the assumption that overt recall responses are diagnostic of the amount and type of covert recall responses (i.e., of the accessibility and quality of clues that come to mind when recall fails). Experiment 3 was designed primarily to test this assumption. Participants were presented with the memory pointers used in Experiment 2, and, when they failed to recall a target, they were asked to list the partial clues that they could retrieve about it (see Lovelace, 1987; Read & Bruce, 1982). It was expected, first, that the amount of partial information accessed would be smaller for the low-ACC pointers than for the CC and CW pointers and, second, that the accuracy of the partial information retrieved would be higher for the CC than for the CW pointers. Such a pattern would explain, for example, why in Experiment 2 low-ACC pointers produced relatively low FOK judgments, whereas the deceptive CW pointers engendered stronger but unwarranted FOK judgments.

Participants in Experiment 3 also made FOK judgments at the end of each trial and were tested on recognition. In this manner, it is also possible to examine how FOK judgments vary with overt accessibility (i.e., with the number of partial clues actually retrieved). It is assumed, however, that participants cannot always spell out all of the clues that come to mind when searching for a solicited target. Therefore, FOK judgments are expected to also correlate with covert accessibility, as indexed by the ACC score.

Finally, an attempt was made in Experiment 3 to identify instances in which the participant's effective target may have been different from the experimenter-defined target (see R. Brown & McNeill, 1966; Koriat & Lieblich, 1974). Therefore, after choosing an answer in the recognition test, participants were asked to indicate whether or not that answer was the one they had been searching for during the retrieval phase.

Method

Participants. Eighteen psychology undergraduates participated in the experiment for course credit.

Stimulus materials. The same 110-item general knowledge test used in Experiment 2 was used here. The recognition booklet was similar to that of Experiment 2 except that the words *intended* and *unintended* appeared after each pair of distractors, whereas the recall booklet was designed to allow filling in both the answer and the specific types of partial information. Thus, the response sheet was laid out in columns labeled as follows: (a) answer, (b) number of syllables, (c) guessed letters (initial, middle, or last), (d) words of similar sound, (e) other clues, (f) late answer, and (g) FOK judgment.

Procedure. The experiment was administered in group sessions lasting about 3 hr; there were 2 to 6 students in each group. The experiment included a recall phase followed by a recognition phase. Students were informed that they would be required to answer a series of general knowledge questions twice, once in an open-ended format, in which the questions would be read by the experimenter (recall phase), and once in a two-alternative forced-choice format (recognition phase). In the recall phase, students were instructed to try to recall the answer. However, they were told that even when they failed to recall it, they sometimes could retrieve partial clues about the solicited name or concept. Thus, they were instructed to write down any such clues that came to mind in the spaces provided on the response sheet. In the column for number of syllables, they could write the number of syllables in the solicited targets. For guessed letters, they could write down the initial letter of the word, any middle letter, the final letter, or a combination of these letter options. For words of similar sound, they could list words having the same sound as the target, and, in the column labeled other clues, they wrote down any additional clues (e.g., those regarding the physical appearance of the person in question or of the place whose name was solicited). Thirty seconds were allotted for each question, at the end of which students supplied the answer if they could recall it at that point and then (and not before) wrote down their FOK judgment on a 50%-100% scale. (The FOK instructions were similar to those of Experiment 2.)

The students were given a short break in the middle of the recall phase. On completion of the recall phase, students were handed the recognition booklet and asked to circle the correct answer. In addition, they were asked to indicate whether that answer was indeed the one they had been looking for in the recall phase ("intended"). When the answer circled was not the one intended, they were required to write down the intended target if they could recall it.

Results

The results are reported in three sections. The first examines the recall and recognition findings that constitute replications of those of Experiment 2. The second focuses on the partial information accessed and examines how the amount and quality of that information differ among the CC, CW, and low-ACC pointers. The third focuses on FOK judgments and examines how they vary with both covert and overt indexes of accessibility.

Recall and recognition performance. On average, students provided an answer to 33.1% of the questions (excluding late answers; see later discussion). This percentage is, of course, much higher than that of Experiment 2 (6.6%). The respective percentages for the CC, CW, and low-ACC pointers were 48.5, 43.2, and 6.8. (The percentages for the CC and CW pointers were also slightly higher than the respective percentages in Experiment 1 [39.7 and 41.4, respectively], even when late answers were not included.) Thus, as expected, the CC and CW pointers precipitated a larger number of (complete) answers than the low-ACC pointers, *F(l,* 17) = 112.74, *p* < .0001. In terms of the quality of the answers, however, the mean OBA was clearly higher for the CC (75.0%) than for the CW (30.0%) pointers, *F{\,* 17) = 119.29,/? < .0001. It should be pointed out that the respective mean for the low-ACC pointers was also very low, 26.4%, suggesting that these pointers as well, to the extent that they elicited an answer at all, were largely deceptive.

A similar pattern was exhibited by the late answers reported, although the percentage of such answers was very low. When initial recall failed, the mean percentages of late answers were 4.2, 5.4, and 0.8 for the CC, CW, and low-ACC pointers, respectively. The respective OBA means were 52.8, 16.8, and 0.0. These latter means were based only on 6 students, 7 students, and 1 student, respectively, for whom the pertinent data were available.

Turning to recognition performance, mean correct recognition percentages for the CC, CW, and low-ACC pointers were 80.3, 51.0, and 46.9, respectively. As would be expected, recognition performance was correlated with recall performance: The mean recognition performance for trials in which a correct answer was recalled was 98.9%, whereas that for trials involving incorrect answers was 34.1% (i.e., a preference for the wrong alternative). Mean recognition for nonrecall trials was 54.9%.

Amount and accuracy of partial information. I now turn to the main aim of Experiment 3: examination of the amount and accuracy of the partial information reported after recall failure. An ACC score was calculated for each of four attributes, depicting the amount of partial information supplied regardless of its correctness. The attributes were letters, syllables, similar sound, and other clues. In the case of letters, the score represented the total number of letters (0-3) supplied in the rubrics initial, middle, and last. The score for syllables was simply 0 or 1, according to whether or not the student reported an estimate of the number of syllables. The scores for similar sound and other clues consisted of the number of phonologically similar strings and the number of other clues reported, respectively. A total score was also calculated, consisting of the number of pieces of information reported; letters, phonological associates, and other clues were each counted as a separate piece of information.

Table 2 presents the means of the accessibility scores for the three classes of pointers (CC, CW, and low-ACC). It was expected that the CC and CW pointers would precipitate a larger amount of partial information than the low-ACC pointers. Indeed, although the overall amount of partial information produced was quite low, the expected pattern was generally found for each of the variables. Analyses comparing the low-ACC pointers with the CC and CW pointers combined yielded the following (one-tailed) results: letters, r(17) = 2.14, p < .05; syllables, *t{\T)* = 2.18, *p < .05;* similar sound, *t(17)* = 1.67, *p* < .10; other clues, f(17) = 4.17, *p* < .001; and total score, f(17) = 7.15, *p* < .0001. Thus, the CC and CW pointers not only elicited a greater proportion of (complete) answers than did the low-ACC pointers but also gave rise to more than twice as many pieces of partial information than the low-ACC pointers when complete recall failed. This pattern may explain the higher FOK judgments associated with omission responses to these pointers in Experiments 1 and 2.

Table 2
Mean Accessibility for the Attributes of Partial Information Tested in Experiment 3 for the Three OBA Classes of Pointers

	OBA class		
Attribute	Consensually correct	Consensually wrong	Low accessibility
Letters	0.122	0.109	0.059
Syllables	0.052	0.059	0.029
Phonology	0.041	0.025	0.017
Other	0.162	0.183	0.069
Total	0.376	0.377	0.174

Note. OBA = output-bound accuracy.

Turning to the quality of the information accessed, an OBA score was calculated for three attributes: letters, syllables, and similar sound. (The information listed under other clues was difficult to score for accuracy. Also, the correct answer for number of syllables was equivocal for four pointers, and these pointers were therefore not scored for the accuracy of number of syllables.) For letters, it consisted of the percentage of correct letters of all of the letters reported by the student. A letter was scored as correct if it appeared in its specific position: first, last, or middle (any letter other than the first or last). The total accuracy score consisted of the percentage of different pieces of information that were correct, excluding those appearing in the other clues category.

Because these scores could be calculated only for students who provided partial information regarding the pertinent attribute, the attribute means (see Table 3) were each based on a small number of students. Nevertheless, these means evidence a trend indicating that the partial information produced was more likely to be correct for the CC pointers than for the CW (or low-ACC) pointers. However, only the

Table 3
Mean Accuracy Scores for the Attributes of Partial Information Tested in Experiment 3 for the Three OBA Classes of Pointers

	OBA class					
	Consensually correct		Consensually wrong		Low accessibility	
Attribute	*M*	*n*	*M*	*n*	*M*	*n*
Letters	39.05	14	27.72	15	27.78	10
Syllables	69.44	9	32.58	11	30.56	9
Phonology	42.86	7	11.11	3	21.43	7
Total accuracy	49.01	16	25.28	16	31.76	14

Note. OBA = output-bound accuracy.

total accuracy score yielded a significant difference between the CC and CW pointers, $t(15) = 2.11$, $p < .05$ (see later analyses).

In sum, the partial information data indicate that pointers classified as high accessibility on the basis of recall performance tend to also precipitate a relatively large amount of partial information when recall of the full answer fails. Furthermore, the accuracy of the partial information retrieved when initial recall fails is also correlated with the correctness of the answers reproduced in commission trials.

FOR judgments and their accuracy. FOK judgments in Experiment 3 were solicited at the end of a trial, about 30 s after the initiation of the search process. Because these judgments were made after the students had had sufficient opportunity for a deliberate and effortful search for the target, they may be assumed to rest not only on a global assessment of the accessibility of partial clues but also on an analytic evaluation involving content considerations (see Koriat, 1993). Therefore, it is of interest to determine whether some of the predictions of the accessibility model still hold for such "late" FOK judgments.

It should be indicated first that, as in the previous experiments, FOK judgments were significantly higher after recall (correct or incorrect, taking into account both initial recall and late answers) than after nonrecall (91.5% and 61.4%, respectively), $F(1, 17) = 361.31$, $p < .0001$. Among recall trials, those in which the answer was correct were associated with higher FOK judgments (95.8%) than those in which the answer was wrong (87.5%), $F(1, 17) = 40.72$, $p < .0001$.

The following analyses are based only on trials in which no complete answer was reported either initially or just before an FOK judgment was made. Two indexes of accessibility for omission responses were available in Experiment 3: covert accessibility, defined operationally in terms of the division of the pointers into the three classes, and overt accessibility, as reflected in the amount of partial information actually reproduced.

With regard to overt accessibility, all trials were divided on the basis of the total score into those in which no partial information was reported (total amount = 0) and those in which some partial information was reported. FOK judgments for these two classes averaged 59.3% and 70.0%, respectively, $F(1, 17) = 58.05$, $p < .0001$, suggesting that FOK increases with the amount of partial information accessed regarding the unrecallable target. This result is in line with previous results (Blake, 1973; Eysenck, 1979; Koriat, 1993; Schacter & Worling, 1985), indicating that FOK judgments increase with the amount of partial information actually retrieved.

The hypothesis was also examined that the accuracy of FOK in predicting recognition performance varies with the accuracy of the partial information reported. However, the analyses yielded little support for this hypothesis, possibly because of the few observations on which they were based.

In terms of covert accessibility, I first focus on the CC and CW pointers for which accessibility estimates were available from Experiment 1. The classification of these pointers as moderate and high ACC (see Experiment 2) was based on the likelihood that a pointer brought to mind a possible answer in Experiment 1. This classification proved effective in predicting the preliminary FOK judgments made in Experiment 2 before students had had a chance to search deliberately for the target. Would this classification also predict the late FOK judgments solicited in Experiment 3, after students had practically exhausted their search for the target and failed to find it? Although the late FOK judgments of Experiment 3 were considerably lower than the preliminary FOK judgments of Experiment 2 (when recall failed), they too revealed the same general pattern, averaging 65.1% and 69.8% for the moderate- and high-ACC pointers, respectively, $F(1, 17) = 10.21$, $p < .01$. [The respective means for FOK judgments made after commission responses were 90.8% and 92.6%, $F(1, 17) = 3.19$, $p < .10$.]

In a second analysis, FOK judgments were compared for the CC, CW, and low-ACC pointers. The averages were 66.9%, 66.4%, and 56.0%, respectively. Thus, CC and CW pointers elicited similar late FOK judgments, but both evoked higher FOK judgments than the low-ACC pointers, $F(1, 17) = 63.42$, $p < .0001$, for CC and $F(1, 17) = 64.09$, $p < .0001$, for CW.

Given the correlation between overt and covert accessibility, it is of interest to examine whether covert accessibility exerts an effect on FOK over and above that of overt accessibility. A two-way, Overt Accessibility (no vs. some partial information accessed) X Covert Accessibility (low-ACC vs. CC and CW combined) ANOVA yielded the following results: overt accessibility, $F(1, 17) = 38.80$, $p < .0001$; covert accessibility, $F(1, 17) = 83.94$, $p < .0001$; and the Overt Accessibility X Covert Accessibility interaction, $F < 1$. The effects of covert accessibility were significant both for pointers yielding no partial information, $F(1, 17) = 54.67$, $p < .0001$, and for pointers yielding partial information, $F(1, 17) = 33.68$, $p < .0001$.

Turning to FOK accuracy, again there was a trend consistent with the pattern observed in Experiment 2: A two-way FOK Level (below median vs. above median) X Class (CC vs. CW vs. low-ACC) ANOVA on recognition performance yielded the following interaction effect: $F(2, 26) = 4.74$, $p < .05$. For the CC class, recognition memory for low-FOK and high-FOK pointers averaged 70.3% and 79.8%, respectively, $F(1, 15) = 4.28$, $p < .10$. For CW pointers, the respective means were 49.8% and 46.6% ($F < 1$); for the low-ACC pointers, these means were 47.0% and 32.5%, $F(1, 14) = 4.08$, $p < .10$. (Note that the pertinent means were not computable for some students, and hence the different sample sizes.) Thus, only for the CC pointers was there a trend indicating better memory performance for high-FOK judgments than for low-FOK judgments; for the CW and low-ACC pointers, the correlation between recognition memory and FOK judgments tended to be negative. In sum, the systematic differences among the three classes of pointers that were observed with regard to preliminary FOK judgments (Experiment 2) were replicated for the late FOK judgments of Experiment 3.

When the participant's target departs from the experimenter's target. Consider, finally, the question of the effective target for FOK judgments (see Koriat, 1994). In most previous studies evaluating the accuracy of FOK, the correct target has served as the effective criterion. This policy is not only consistent with the spirit of the trace-access model of FOK but is a sensible choice in the absence of direct information regarding the participant's actual target. An exception to this policy, however, can be found in the study of the TOT state (e.g., see R. Brown & McNeill, 1966; Koriat & Lieblich, 1974). Here the accuracy of the partial information reported by the participant regarding unrecallable targets is often evaluated against the participant's declared target when it deviates from the correct (experimenter's) target (and when the participant succeeds in retrieving his or her own target). In the present study, students were specifically instructed to indicate whether the answer they chose in the recognition phase was indeed the one they had been searching for during the recall phase.

These data allowed a two-way classification of trials in terms of whether the answer chosen was correct or not and whether that answer was intended or unintended. Only trials in which no answer had been recalled were included in the following analyses. Overall, when the correct answer was chosen in the recognition test, that answer was marked as intended in 57.0% of the cases across all students and pointers. This percentage, as may be expected, differed for the three classes of pointers (81.1% for the CC pointers, 42.7% for the CW pointers, and 18.2% for the low-ACC pointers). On the other hand, when the incorrect distractor was chosen, that distractor was marked as intended in 53.0% of the cases overall. The respective percentages were 66.7% for the CC pointers, 53.9% for the CW pointers, and 47.5% for the low-ACC pointers. Thus, the correct answer was more likely to be the intended target for the CC pointers than the incorrect distractor, whereas the reverse was true for the CW and low-ACC pointers. Note, however, that a sizable percentage of the correct answers in the CC class were nevertheless judged as unintended (18.9), and an even larger percentage of the incorrect answers in the CW class (46.1) were judged as unintended.

Recall that when students marked their chosen recognition alternative as unintended, they were urged to recall their effective target. These data are not considered here because there were too few instances in which students reported a target other than the one they had already reported in the recall phase.

Consider now the accuracy of the partial information retrieved. The previously reported analyses implicitly assumed that the student's effective target was the same as the experimenter-defined target. In the following analyses, the possibility was examined that the accuracy of the partial information was actually better when the analysis was confined to trials in which the correct target was also the one intended by the student. Indeed, when the correct answer was chosen in the recognition test, the mean total accuracy score was considerably higher when that answer was intended (70.8; $n = 14$) than when it was unintended (23.9; $n = 15$), $r(12) = 3.49$, $p < .005$. In contrast, when the *incorrect* distractor was endorsed, the mean total accuracy score (which was calculated with the experimenter's target as the criterion) was low whether the answer was intended (30.5; $n = 11$) or unintended (14.9; $n = 16$), $r(10) = 1.66$, *ns*.

The incorrect distractor can be seen to represent the student's effective target when that distractor was checked as the one intended. There were only 19 instances overall (across 11 students) in which a student failed to recall the answer, provided partial information about it, chose the incorrect distractor in the recognition test, and checked that distractor as the one intended. For these cases, the total accuracy score, calculated with the student's target as the criterion, averaged 50.8, in comparison with the 30.5 value reported earlier when the experimenter's target was used as the criterion, $f(10) = 2.78$, $p < .01$. It would seem, then, that the partial information provided by participants is generally more accurate when judged against the participants' effective target than when judged against the experimenter's target.

Discussion

The results of Experiment 3 help explain the two major findings of Experiments 1 and 2. First, they help clarify why the classification of memory pointers in terms of the likelihood of precipitating a full answer is effective in predicting FOK judgments even when recall fails. The results indicate that pointers that yield a large number of recall responses overall tend to also precipitate a large amount of partial information when recall fails. If FOK judgments are affected by the amount of partial information retrieved about a momentarily inaccessible target, then pointers eliciting many answers would be expected to evoke a stronger FOK after recall failure than pointers eliciting fewer answers.

Despite the positive correlation between the number of answers precipitated by a pointer (Experiment 1) and the amount of partial information it tends to elicit in omission trials (Experiment 3), these two indexes were found to make independent contributions to FOK judgments in Experiment 3. Possibly, each of these two measures of accessibility represents only a rough index of the kind of partial information assumed to underlie FOK judgments.

Second, the results of Experiment 3 help explain the finding that the quality of the (complete) answers elicited by a pointer in commission trials is also predictive of the accuracy of FOK in omission trials. Pointers producing a large proportion of correct recalls (in Experiment 1) were found to produce a large proportion of correct partial information when recall failed (Experiment 3). Such pointers, according to the accessibility model, should also yield a high FOK accuracy, as was indeed the case in Experiments 1 and 2.

In addition to clarifying the findings of the previous experiments, the results of Experiment 3 also demonstrate the continuity between preliminary FOK (Experiment 2) and late FOK (Experiment 3). Thus, the systematic differences among the three classes of pointers that were observed with regard to the preliminary FOK judgments of

Experiment 2 persisted for the late FOK judgments elicited in Experiment 3. These results testify for the reliability of interpointer differences in properties that are pertinent to both FOK level and FOK accuracy (as discussed subsequently).

General Discussion

The results of the present study provide further evidence in support of the accessibility model of FOK proposed by Koriat (1993). Before summarizing and discussing these results, I comment on some of the unique properties of the methodology adopted in the present study.

Methodological Considerations

A proper test of the accessibility model requires an assessment of the amount and accuracy of the partial information that comes to mind when recall of the complete answer fails. This can be achieved relatively easily with newly acquired artificial stimuli such as those used by Blake (1973), Schacter and Worling (1985), and Koriat (1993). For example, one of the tasks used by Koriat (1993, Experiment 1) required the memorization of experimentally presented four-letter nonsense strings. This task provided a simple measure of accessibility and accuracy in terms of the number of correct and incorrect letters recalled. The results indicated that FOK judgments did indeed increase with the number of letters recalled (both correct and incorrect). They also showed that the accuracy of the FOK derived primarily from the accuracy of the partial information retrieved.

Although such memory tasks are convenient for testing the main propositions of the accessibility model, they have several serious limitations. First, they are not representative of situations in which FOK is typically experienced in everyday life. A strong FOK is often experienced in connection with the retrieval of real-world knowledge, particularly the retrieval of names and concepts from long-term memory (e.g., R. Brown & McNeill, 1966; Cohen, 1990; see Koriat, 1994). Thus, an important goal of the present study was to extend investigation of the accessibility model to such situations.

A second limitation is that such episodic memory tasks provide little opportunity for studying what happens when memory is truly distorted. Only under very circumscribed conditions does memory for episodic information produce a large enough ratio of incorrect to correct recalls (e.g., Jacoby & Whitehouse, 1989; Loftus, 1977; Weingardt et al., 1994). On the other hand, long-term memory for real-world knowledge exhibits many occasions in which, for one reason or another (see Fischhoff et al., 1977), recall or recognition tends to deviate reliably from veridicality. A dissociation between objective and subjective knowing was demonstrated in the present study by exploiting such occasions (see also Koriat, 1976).

Finally, one weakness of the methodology used by Koriat (1993) is that accessibility was measured in terms of the amount of information actually provided by the participant before making FOK judgments. This raises the possibility that the correlation between FOK and the amount of partial information reproduced was due to demand characteristics: Participants felt committed to making stronger FOK judgments after having produced more partial information than after having produced less such information. In the present study, this methodological problem was circumvented by a focus on covert accessibility.

Although the methodology of the present study helps to overcome some of these limitations, it rests on two important assumptions that need to be made explicit. First, it assumes that memory pointers differ reliably across participants, both in the amount of information they tend to precipitate and in the quality of that information. This assumption is supported by the results of Koriat and Lieblich (1977), who found, for example, that memory pointers differed reliably in the extent to which they evoked a TOT state across participants. The results of the present study also lend further support to this assumption and indicate, as well, that pointer characteristics pertinent to FOK are relatively stable across different testing conditions (see later discussion).

The second assumption is that normative data, averaged over participants, provide information about processes that take place within participants (e.g., see Schreiber & Nelson, 1994): Memory pointers that elicit an answer among many participants are assumed to also produce many answers, perhaps covertly, in each individual participant. Furthermore, such pointers are assumed to leave behind more partial activations when the search of the target fails than pointers eliciting an answer among fewer participants. This assumption, in its strongest form, is clearly untenable. After all, there are vast individual differences in participants' familiarity with different topics and in the specific semantic and episodic information they bring to bear on each question. For example, Nelson, Leonesio, Landwehr, and Narens (1986) found individuals' own FOK ratings to be more accurate in predicting recognition performance than normative FOK ratings based on group averages, suggesting that participants have idiosyncratic information that they can use in making FOK judgments (see also Jameson, Nelson, Leonesio, & Narens, 1993).

Nevertheless, the results on the whole do lend credence to the idea that normative data about pointers can provide a rough estimate of the amount and quality of information actually accessible to each individual participant. In fact, these results testify for the general usefulness of the pointer-based methodology.

I now review and discuss the main findings of the present study with regard to the basis of FOK and the reasons for its accuracy. The first section examines the evidence supporting the claim that FOK judgments monitor the accessibility of information pertaining to a solicited target, regardless of its accuracy. The next two sections take up the question of FOK accuracy and examine the two factors assumed to contribute to the predictive validity of FOK: the stability of interpointer differences in both accessibility and accuracy and the correlation between accessibility and accuracy. The final section examines the implications of the present study with regard to the link between memory and metamemory.

Effects of Accessibility on FOK

Consider first the question of how the feeling of knowing is formed. The results, on the whole, support the proposition that FOK judgments monitor the mere accessibility of information pertaining to the target, regardless of the correctness of that information. There are two lines of evidence in support of this proposition. The first concerns overt accessibility. In this study, consistent with previous findings (Krinsky & Nelson, 1985; Nelson & Narens, 1990), FOK judgments were markedly higher after commission responses than after omission responses. This was true both when the commission answer was correct and when it was wrong, a pattern observed in all three experiments. Thus, in Experiment 1, FOK judgments for commission responses averaged 90.5%, in comparison with only 59.1% for omission responses. The respective means for Experiment 2 were 98.3% and 70.0%, and those for Experiment 3 were 91.5% and 61.4%. It appears, then, that the mere accessibility of a full answer constitutes strong subjective evidence that one "knows" the correct answer.

It should be noted, however, that FOK judgments were nevertheless higher after correct commissions than after incorrect commissions. The respective means were 94.8% and 84.2% in Experiment 1, 99.1% and 98.0% in Experiment 2, and 95.8% and 87.5% in Experiment 3. This difference could derive from the effects of content considerations (see Koriat, 1993). However, it could also stem from the use of cues pertaining to the intensity of the information retrieved, particularly its ease of access. Indeed, although the CC and CW pointers yielded similar rates of commission responses in Experiments 1 and 3, the CC pointers produced a larger percentage of answers (12.3) than the CW pointers (7.2) under the speeded instructions of Experiment 2, $F(1, 29) = 7.33$, $p < .05$. This pattern suggests that the answers to the CC pointers have a shorter latency of recall than the CW answers. Furthermore, in consideration only of the matched CC and CW pointers, FOK judgments for these pointers after a commission response were higher in the speeded recall task of Experiment 2 (98.6%) than in Experiment 1 (88.3%), presumably because the former responses included a higher proportion of easily retrieved answers. Taken together, these results suggest that correct responses are retrieved faster than incorrect responses and that ease of access also affects FOK, further enhancing its accuracy. Indeed, Koriat (1993) found evidence suggesting that participants can monitor the accuracy of information retrieved on the basis of ease of access, as indexed by recall latency. Other researchers, too, have viewed ease of access as an important cue for judgments of learning (Begg et al., 1989), subjective confidence (Kelley & Lindsay, 1993), and sense of familiarity (e.g., Jacoby & Kelley, 1987).

Another demonstration of the dependence of FOK on overt accessibility comes from the partial information results of Experiment 3. These results indicate that, when recall of the full answer is aborted, FOK is higher when partial information is accessed about the solicited target (70.0%) than when no partial information is retrieved (59.3%).

The second line of evidence pertains to covert accessibility, as inferred from the percentage of commission responses precipitated by a pointer across students in Experiment 1. Pointers engendering many answers were associated with higher FOK judgments than pointers engendering few answers, and this held true even when overt accessibility was controlled. Thus, high-ACC pointers were associated with higher FOKjudgments both among students who provided a full answer (Experiments 1 and 2) and among those who failed to come up with an answer (Experiments 1, 2, and 3). Furthermore, in those cases in which students failed to produce an answer in Experiment 3, high-ACC pointers yielded higher FOK judgments than low-ACC pointers, both among those who recalled partial information and among those who failed to recall any partial information.

In terms of omission responses, these results suggest that even when recall is aborted, high-ACC pointers leave behind more partial clues than low-ACC pointers. The results of Experiment 3 support this idea: The amount of partial information reproduced after recall failure was higher for the high-ACC pointers than for the low-ACC pointers. It should be noted, however, that differences in inferred, covert accessibility exerted an effect on FOK judgments over and above the effects of overt accessibility, suggesting that the covert accessibility measure captures aspects of partial information that were not tapped in Experiment 3.

With regard to commission responses, the effects of covert accessibility on FOKjudgments may be due to systematic differences between low-ACC and high-ACC pointers in the intensity of the recalled answer (e.g., its ease of access, vividness, specificity, and persistence; see Koriat, 1993). Another possibility is that FOK is stronger when a pointer brings to mind many answers in addition to the one chosen than when it evokes just one answer. If this latter possibility is correct, it would support the contention (see Koriat, 1993, 1994) that FOK depends on a nonanalytic process that considers the mere amount of information accessible without regard to the agreement or conflict between the various pieces of information accessible. Indeed, in Experiment 2, there was no difference in average FOK judgments between deceptive items that evoked few and those that evoked many different incorrect answers across students in Experiment 1. If FOK were determined by a calculated inference, then an answer selected from a small set of candidates would have been expected to result in a stronger FOK than one selected from many candidates (but see Schreiber & Nelson, 1994). Thus, FOK judgments, particularly those solicited early in the search process, may be assumed to respond to the features rather than to the conjunctions of the activated information (to borrow Treisman's distinction [e.g., Treisman & Gelade, 1980]). Only when the search becomes more deliberate and analytic would FOK be expected to also depend on the compatibility between the various clues that come to mind.

In sum, the results of the present study concur with those of previous studies (Blake, 1973; Brown & Bradley, 1985; Eysenck, 1979; Koriat, 1993; Lories, in press; Metcalfe et al., 1993; Schacter & Worling, 1985; Schwartz & Smith, 1994) in supporting the hypothesis that FOK judgments monitor the accessibility of information pertaining to the solicited target. These results were obtained for a general information memory task tapping long-term memory. Furthermore, they were obtained even under conditions that did not require report of the retrieved partial clues, thus eliminating an explanation of the accessibility-FOK correlation in terms of the effects of demand characteristics.

The effects of accessibility on FOK judgments were found regardless of the (assumed or tested) correctness of the accessible information. Commission responses were associated with higher FOK judgments than omission responses when the answer was correct as well as when it was wrong, and the effects of accessibility were found for both CC and CW pointers. In fact, covert accessibility affected FOK without affecting actual recognition performance. Thus, FOK seems to monitor primarily the mere accessibility of information.

Despite the reliable effects of accessibility documented in all three experiments, it is important to note that these effects leave a large amount of the variance of FOK still to be explained. It can be seen, for example, that the low-ACC pointers yielded a wide range of FOK judgments (Figure 5). Part of the unexplained variance may be due to the crude nature of the measure of accessibility used, particularly the fact that it was based on normative data that did not capture some of the idiosyncratic variance. Another part, however, may reflect the influence of other factors that affect FOK (see Nelson et al., 1984).

Stability of Accessibility Indexes of Pointers

I now examine the basis for the general accuracy of FOK in predicting memory performance. A necessary condition for the predictive validity of FOK is the existence of reliable interpointer differences in properties that are pertinent to both FOK judgments and actual memory performance. In the absence of such differences, the FOK-recognition correlation would be very close to zero (see Koriat, 1993; Nelson & Narens, 1990; Schwartz & Metcalfe, 1994). Assuming that FOK judgments monitor primarily the quantity of information accessible, the predictive validity of these judgments should depend specifically on the extent to which interpointer differences in the amount of information accessible at Time 1 correlate with interpointer differences in the correctness of memory answers at Time 2 (see Koriat, 1993, 1994). This correlation, in turn, should depend on two factors: (a) the stability of interpointer differences (over time and across testing conditions) in both the quantity and accuracy of the information they elicit and (b) the correlation between accessibility and accuracy (i.e., the extent to which the mere quantity of information accessible about a target is diagnostic of the amount of correct information accessible about it). In this section, I consider the former factor, leaving discussion of the latter one to the next section.

The results of the present study testify for the stability of interpointer differences in both the likelihood of precipitating complete or partial recall and the likelihood of selectively evoking correct rather than incorrect recalls.

As far as the former is concerned, pointers eliciting an answer rate of less than 37% in Experiment 1 (low ACC) and those eliciting a rate of more than 37% (high ACC) produced rates of 5.7% and 14.0%, respectively, in the speeded recall task of Experiment 2, $t(29) = 4.01$, $p < .0005$. These pointers also differed, in Experiment 3, in both initial recall (33.6% and 58.9%, respectively), ?(17) = 7.68, $p < .0001$, and late recall (after an initial recall failure; 2.2% and 10.2%, respectively), f(17) = 1.83, $p < .05$ (one-tailed).

A similar pattern emerged when the three classes of pointers (CC, CW, and low-ACC) were compared across Experiments 2 and 3. The former two classes yielded a higher percentage of recall than the last class in both Experiment 2 (9.7 and 0.3, respectively) and Experiment 3 (45.9 and 6.8, respectively). They also yielded a higher percentage of late answers after an initial recall failure (4.9 and 0.8, respectively) in Experiment 3.

Of particular interest is the relationship between complete and partial recall: Although the CC and CW pointers elicited substantially more (full) answers than the low-ACC pointers in Experiment 3, they also induced more partial information than the low-ACC pointers when recall of the complete answer failed. Furthermore, among the CC and CW pointers, those that elicited more answers in Experiment 1 also elicited more partial information after recall failure in Experiment 3.

Thus, pointers differ reliably in the extent to which they tend to bring to mind full or partial answers, and these differences were relatively stable over the time interval represented in the three experiments, which ranged from 5 s to 30 s. Although not investig'ated here, it may be conjectured that pointers that bring to mind more partial clues very early in the search process are also more likely to produce complete recall later. This correlation constitutes a basic ingredient in the predictive validity of FOK after initial recall failure.

Turning to the accuracy of the information reproduced, this too appears to constitute a stable characteristic of pointers. Pointers classified as CC and CW in Experiment 1 also differed in the correctness of the answers produced in both the speeded recall task of Experiment 2 and the extended recall task of Experiment 3. In the latter experiment, they also differed in the correctness of the late answers produced when initial recall failed.

A positive correlation was also found between the quality of the full answers produced (in Experiment 1) and the quality of the partial information recovered (in Experiment 3): Pointers engendering accurate recall also precipitated accurate partial recall when complete recall failed. In addition, there was a strong correlation between recall and recognition performance in all three experiments: Pointers producing predominantly correct recall yielded substantially better recognition performance than pointers producing predominantly incorrect recall (see Figures 2, 4, and 5).

In sum, differences between memory pointers are relatively stable in both of the characteristics that are pertinent to FOK: the likelihood of producing complete or partial recall and the accuracy of the information that they bring to mind. Both of these constitute third-person characteristics (see Nelson et al., 1986) whose stability provides the necessary grounds for the general validity of a first-person variable (FOK). It is assumed that participants have direct access to the amount of information recalled but not to the accuracy of that information.

Validity of Accessibility-Based Predictions

Consider next the second determinant of FOK accuracy: the correlation between accessibility and accuracy. It was proposed that the accuracy of FOK stems from the accuracy of memory itself (i.e., from the fact that the information retrieved from memory is generally correct). Therefore, pointers differ primarily between those that elicit little information and those that produce a large amount of correct information. Under such conditions (and given the stability of interpointer differences in the quantity and accuracy of the information they tend to precipitate), the accessibility of information at one point in time should predict the likelihood of recalling or identifying the correct target at a later point in time.

This is indeed what happened for the CC pointers in the present study, which were representative of memory pointers in general. Thus, for the sample of 37 CC pointers used in Experiments 2 and 3, there was a .92 correlation in Experiment 1 between the total number of answers precipitated by a pointer and the total number of correct answers reported. The correlation between the former and the number of incorrect answers reproduced was only .46. This pattern of correlations can explain part of the predictive validity of FOK. Indeed, in Experiment 1, the total number of answers elicited by a pointer correlated .47 ($p < .005$) with the percentage of correct recognition for that pointer across the 37 CC pointers. Thus, the mere accessibility of information at one point in time is predictive of correct recognition at a later point. As expected, in all three experiments, positive within-subject correlations were found between FOK judgments and subsequent recognition performance.

CW pointers, on the other hand, are atypical and were deliberately chosen to destroy the common predominance of correct recalls over commission errors and to produce a dissociation between objective and subjective indexes of knowing (see also Koriat, 1976). Indeed, for the 37 CW pointers, the total number of answers reproduced correlated .43 with the number of correct answers and .94 with the number of incorrect answers. For this class of pointers, the total number of answers precipitated by a pointer correlated — .40 ($p < .05$) with percentage of correct recognition. Thus, accessibility here is predictive of false recognition.

As expected, the CW pointers produced two findings. First, in all three experiments, FOK judgments were overly inflated in comparison with actual recognition performance. The discrepancy was much more extreme than the general overconfidence typically observed in calibration studies (see Koriat, Lichtenstein, & Fischhoff, 1980) and more similar to that reported by Fischhoff et al. (1977). Second, there was little within-subject correlation between FOK judgments and actual recognition memory in Experiments 1 and 3, and the correlation in Experiment 2 was, in fact, negative. This negative correlation was obtained even though the analysis was confined to omission responses, possibly because, for the CW pointers, accessibility is predictive of a higher rate of incorrect partial information when recall fails (Experiment 3).

These results clearly indicate that FOK accuracy is not inherent to memory functioning, as is implied by the trace-access view, but is highly dependent on the specific properties of the sampled items, particularly the overall correctness of the information that comes to mind. In this sense, the accuracy of metamemory can be said to depend on the accuracy of memory itself (see Perfect & Stollery, 1993).

Relationship Between Memory and Metamemory

The results of the present study have some bearing on the general relationship between memory and metamemory and, in particular, on the relationship between the processes of memory retrieval and memory monitoring. The trace-access view of FOK (e.g., Hart, 1965; Yaniv & Meyer, 1987) generally maintains a modular organization of monitoring and retrieval, with the former preceding the latter (see Koriat, 1994). A similar conception is implied in some of the discussions of the cue-familiarity account (e.g., Reder & Ritter, 1992). These discussions assume that people are able to make FOK judgments before actually attempting to answer a question, suggesting that such judgments may come from a processing stage that is independent of the retrieval process itself. Other authors, still, have proposed a distinction between two types of FOK judgments, a preliminary FOK that occurs before the initiation of a deliberate search for the target and an FOK that occurs after a retrieval attempt has failed (see Barnes et al., 1994; Schwartz, 1994).

The accessibility view of FOK, in contrast, assumes an interactive process in which monitoring and retrieval operations are interwoven from the start, each feeding into the other (Koriat, 1993, 1994): It is by attempting to search for the solicited target that one can assess the likelihood that the target will be recalled or recognized in the future. Indeed, the results of the present study motivate a more continuous view of FOK in which FOK judgments are computed and updated on line on the basis of the clues accumulated at each point in time. Thus, interpointer differences in accessibility, defined in terms of the likelihood of precipitating an answer in Experiment 1, were found to predict FOK judgments solicited at different stages of the search for the target. These judgments included the speeded, preliminary FOK judgments solicited before students had had sufficient time for a deliberate search (Experiment 2), the "typical"

FOK judgments collected after students had given up searching for the target (Experiment 1), and the late FOK judgments of Experiment 3, which were solicited after a 30-s interval during which the student attempted to retrieve partial information. In addition, the CC and CW pointers, which were associated with considerably higher preliminary FOK judgments in Experiment 2 than the low-ACC pointers (Figure 4), also produced stronger late FOK judgments in Experiment 3 than the low-ACC pointers.

The stability of interpointer differences was demonstrated in the present study not only for FOK level but for FOK accuracy: Pointers precipitating mostly correct answers in Experiment 1 also yielded more accurate FOK judgments than those precipitating mostly incorrect answers. This difference in FOK accuracy was also observed for the preliminary FOK judgments of Experiment 2, for FOK judgments elicited after recall failure in Experiment 1, and for the late FOK judgments of Experiment 3. These results are difficult to explain in terms of the idea that FOK depends on a preretrieval stage.

The results demonstrate a certain degree of continuity in FOK judgments across the retrieval phases sampled in the present study and support a conceptualization in which monitoring and retrieval are seen to interact from the start, with FOK judgments being based on the products of retrieval (Koriat, 1993, 1994). Thus, there is little advantage in postulating a separate, presearch monitoring mechanism because even the early, preliminary FOK is based on an initial, rudimentary search (not necessarily through candidate answers) and on assessment of the overall accessibility of information that comes to mind.

It is important to note, however, that the relative stability of FOK judgments over the different phases of the retrieval process does not deny the possibility of systematic interpointer differences in the rate of information accumulation over time. In fact, such differences have been postulated to account for some of the discrepancies observed between FOK and subsequent memory performance (Koriat, 1993, 1994; Koriat & Lieblich, 1977; see also Morris, 1990). As the memory search changes from a shallow analysis that considers diffuse activations emanating from a broad memory region to a more focused and detailed analysis, different features of each pointer can achieve prominence. Therefore, pointers may differ not only in the overall level of FOK judgments they produce but in the way in which such judgments change over the course of the retrieval process.

The idea that monitoring processes rest on the output of search and retrieval processes implies that metamemory judgments actually mirror memory processes and have no privileged access to information other than that which is accessible to the retrieval process. This idea has broad implications that should be explored in future research. First, it bears on the general issue of how people validate their memories (see Brewer, in press; Ross, in press). Second, it helps specify some of the reasons for the accuracy and inaccuracy of metacognitive judgments (see Schwartz, 1994). In particular, it can help clarify the conditions expected to lead to an "illusion of knowing" (Glenberg, Wilkinson, & Epstein, 1982). Finally, it may have implications regarding the presumed causal link between metamemory and memory in cognitive development (see Schneider, 1985).

References

Barnes, A. E., Nelson, T. O., Dunlosky, J., Mazzoni, G., & Narens, L. (1994). *An integrative system of metamemory components involved in retrieval.* Manuscript submitted for publication.

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language, 28,* 610-632.

Blake, M. (1973). Prediction of recognition when recall fails: Exploring the feeling-of-knowing phenomenon. *Journal of Verbal Learning and Verbal Behavior, 12,* 311-319.

Brewer, W. F. (in press). What is recollective memory? In D. C. Rubin (Ed.), *Remembering our past: Studies in autobiographical memory.* Cambridge, England: Cambridge University Press.

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin, 109,* 204-223.

Brown, A. S., & Bradley, C. K. (1985). Semantic prime inhibition and memory monitoring. *Bulletin of the Psychonomic Society, 23,* 98-100.

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5,* 325-337.

Cohen, G. (1990). Why is it difficult to put names to faces? *British Journal of Psychology, 81,* 287-297.

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, and Cognition, 18,* 142-150.

Durso, F. T., & Shore, W. J. (1991). Partial knowledge of word meanings. *Journal of Experimental Psychology: General, 120,* 190-202.

Eysenck, M. W. (1979). The feeling of knowing a word's meaning. *British Journal of Psychology, 70,* 243-251.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 552-564.

Gardiner, J. M., Craik, F. I. M., & Bleasdale, F. A. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition, 1,* 213-216.

Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition, 15,* 84-93.

Glenberg, A. M., Wilkinson, A. C, & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition, 10,* 597-602.

Gruneberg, M. M., & Monks, J. (1974). Feeling of knowing and cued recall. *Ada Psychologica, 38,* 257-265.

Gruneberg, M. M., Monks, J., & Sykes, R. N. (1977). Some methodological problems with feeling of knowing studies. *Acta Psychologica, 41,* 365-371.

Gruneberg, M. M., Smith, R. L., & Winfrow, P. (1973). An investigation into response blockaging. *Acta Psychologica, 37,* 187-196.

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

Hart, J. T. (1967a). Memory and the memory-monitoring process. *Journal of Verbal Learning and Verbal Behavior, 6,* 685-691.

Hart, J. T. (1967b). Second-try recall, recognition and the memory-monitoring process. *Journal of Educational Psychology, 58,* 193-197.

Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception, and concept learning. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 18, pp. 1-47). San Diego, CA: Academic Press.

Jacoby, L. L., & Kelley, C. M. (1987). Unconscious influences of memory for a prior event. *Personality and Social Psychology Bulletin, 13,* 314-336.

Jacoby, L. L., & Kelley, C. M. (1991). Unconscious influences of memory: Dissociations and automaticity. In D. Milner & M. Rugg (Eds.), *The neuropsychology of consciousness* (pp. 201-233). San Diego, CA: Academic Press.

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General, 118,* 126-135.

Jacoby, L. L., Woloshyn, V., & Kelley, C. M. (1989). Becoming famous without being recognized: Unconscious influences of memory provided by divided attention. *Journal of Experimental Psychology: General, 118,* 115-125.

Jameson, A., Nelson, T. O., Leonesio, J., & Narens, L. (1993). The feeling of another person's knowing. *Journal of Memory and Language, 32,* 320-335.

Jameson, K. A., Narens, L., Goldfarb, K., & Nelson, T. O. (1990). The influence of subthreshold priming on metamemory and recall. *Acta Psychologica, 73,* 55-68.

Jones, G. V. (1989). Back to Woodworth: Role of interlopers in the tip-of-the-tongue phenomenon. *Memory & Cognition, 17,* 69-76.

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32,* 1-24.

Koriat, A. (1975). Phonetic symbolism and feeling of knowing. *Memory & Cognition, 3,* 545-548.

Koriat, A. (1976). Another look at the relationship between phonetic symbolism and the feeling of knowing. *Memory & Cognition, 4,* 244-248.

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. (1994). Memory's knowledge of its own knowledge: The accessibility account of the feeling of knowing. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 115-135). Cambridge, MA: MIT Press.

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General, 123,* 297—315.

Koriat, A., & Goldsmith, M. (in press). Memory as something that can be counted vs. memory as something that can be counted on. In D. Herrman, M. Johnson, C. McEvoy, C. Hertzog, & P. Hertel (Eds.), *Basic and applied memory: Research on practical aspects of memory.* Hillsdale, NJ: Erlbaum.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 107-118.

Koriat, A., & Lieblich, I. (1974). What does a person in a "TOT" state know that a person in a "don't know" state doesn't know? *Memory & Cognition, 2,* 647-655.

Koriat, A., & Lieblich, I. (1977). A study of memory pointers. *Acta Psychologica, 41,* 151-164.

Krinsky, R., & Nelson, T. O. (1985). The feeling of knowing for different types of retrieval failure. *Acta Psychologica, 58,* 141-158.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306-334). Cambridge, England: Cambridge University Press.

Loftus, E. F. (1977). Shifting human color memory. *Memory & Cognition, 5,* 696-699.

Lories, G. (in press). *Partial knowledge, distractor elimination and FOK accuracy.* Manuscript submitted for publication.

Lovelace, E. (1987). Attributes that come to mind in the TOT state. *Bulletin of the Psychonomic Society, 25,* 370-372.

Metcalfe, J. (1993). Novelty monitoring, metacognition, and control in a composite holographic associative recall model: Implications for Korsakoff amnesia. *Psychological Review, 100,* 3-22.

Metcalfe, J. (1994). A computational modeling approach to novelty monitoring, metacognition, and frontal lobe dysfunction. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 137-156). Cambridge, MA: MIT Press.

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 851-861.

Metcalfe, J., & Shimamura, A. P. (1994). *Metacognition: Knowing about knowing.* Cambridge, MA: MIT Press.

Miner, A. C, & Reder, L. M. (1994). A new look at feeling of knowing: Its metacognitive role in regulating question answering. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 47-70). Cambridge, MA: MIT Press.

Morris, C. C. (1990). Retrieval processes underlying confidence in comprehension judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 223-232.

Narens, L., Jameson, K. A., & Lee, V. A. (1994). Subthreshold priming and memory monitoring. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 71-92). Cambridge, MA: MIT Press.

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 judgment for predicting perceptual identification and relearning. *Journal of Experimental Psychology: General, 113,* 282-300.

Nelson, T. O., Leonesio, R. J., Landwehr, R. F., & Narens, L. (1986). A comparison of three predictors of an individual's memory performance: The individual's feeling of knowing versus the normative feeling of knowing versus base-rate item difficulty. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 279-287.

Nelson, T. O., & Narens, L. (1980). A new technique for investigating the feeling of knowing. *Acta Psychologica, 46,* 69-80.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 26, pp. 125-173). San Diego, CA: Academic Press.

Perfect, T. J., & Stollery, B. (1993). Memory and metamemory performance in older adults: The one deficit or two? *Quarterly Journal of Experimental Psychology, 46A,* 119-135.

Read, J. D., & Bruce, D. (1982). Longitudinal tracking of difficult memory retrievals. *Cognitive Psychology, 14,* 280-300.

Reason, J., & Lucas, D. (1984). Using cognitive diaries to investigate naturally occurring memory blocks. In J. E. Harris & P. E. Morris (Eds.), *Everyday memory, actions and absentmindedness* (pp. 53-69). San Diego, CA: Academic Press.

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology, 19,* 90-138.
Reder, L. M. (1988). Strategic control of retrieval strategies. *Psychology of Learning and Motivation, 22,* 227-259.
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, and Cognition, 18,* 435-451.
Ross, M. (in press). Validating memories. In N. L. Stein, P. A. Ornstein, B. Tversky, & C. Brainerd (Eds.), *Memory for everyday and emotional events.* Hillsdale, NJ: Erlbaum.
Schacter, D. L., & Worling, J. R. (1985). Attribute information and the feeling of knowing. *Canadian Journal of Psychology, 39,* 467-475.
Schneider, W. (1985). Developmental trends in the metamemory-memory behavior relationship: An integrative review. In D. L. Forrest-Pressley, G. E. MacKinnon, & T. G. Waller (Eds.), *Cognition, metacognition, and human performance* (Vol. 1, pp. 57-109). New York: Academic Press.
Schreiber, T. A., & Nelson, D. L. (1994). *Feelings of knowing and retrieval processes.* Manuscript submitted for publication.
Schwartz, B. L. (1994). Sources of information in metamemory: Judgments of learning and feelings of knowing. *Psychonomic Bulletin and Review, 1,* 357-375.
Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1074-1083.
Schwartz, B. L., & Metcalfe, J. (1994). Methodological problems and pitfalls in the study of human metacognition. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 93-113). Cambridge, MA: MIT Press.
Schwartz, B. L., & Smith, S. M. (1994). *The products of retrieval and cue familiarity influence tip-of-the-tongue states.* Manuscript submitted for publication.
Shimizu, H., & Kawaguchi, J. (1993). The accuracy of feeling-of-knowing judgments for general-information questions using the recall retest method. *Japanese Psychological Research, 35,* 215-220.
Smith, S. M. (1994). Frustrated feelings of imminent recall: On the tip of the tongue. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 27-45). Cambridge, MA: MIT Press.
Treisman, A. M., & Gelade, G. A. (1980). Feature integration theory of attention. *Cognitive Psychology, 12,* 97-136.
Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5,* 381-391.
Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 4,* 207-232.
Weingardt, K. R., Leonesio, R. J., & Loftus, E. F. (1994). Viewing eyewitness research from a metacognitive perspective. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 157-184). Cambridge, MA: MIT Press.
Yaniv, I., & Meyer, D. E. (1987). Activation and metacognition of inaccessible stored information: Potential bases for incubation effects in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 187-205.

Received October 11, 1994
Revision received March 2, 1995
Accepted March 28, 1995 ■