

The Cue-Familiarity Heuristic in Metacognition

Janet Metcalfe, Bennett L. Schwartz, and Scott G. Joaquim

Four experiments contrasted the cue-familiarity hypothesis of feeling-of-knowing judgments (FKJs) and tip-of-the-tongue feelings (TOTs) to the target-retrievability hypothesis. Familiarity of the cues was contrasted to memorability of the targets in a paired-associate design (e.g., A-B A-B, A-B A-B', A-B A-D, A-B C-D), in which the number of repetitions of the cue A terms was dissociated from the memorability of the target B terms. Little support was found for the target-retrievability hypothesis, because in none of the 4 experiments were FKJs related to target memorability. In one experiment, an omnibus retrieval hypothesis (which implicates total retrieval rather than just correct retrieval) and the cue-familiarity hypothesis produced isomorphic predictions that were borne out by the FKJ and TOT results. All 4 experiments supported the cue-familiarity hypothesis, because FKJs and TOTs were directly related to the number of presentations (and thereby the familiarity) of the cues.

This article investigates how people make judgments about their own future cognition. Much research has been directed at the manner in which people make judgments under uncertainty (Kahneman, Slovic, & Tversky, 1982; Kahneman & Tversky, 1973; Tversky & Kahneman, 1973). Traditionally, this research has focused on predictions about events in the world, rather than internally directed predictions. And there has been some progress in understanding the heuristics people use. Recently, though, there has been an upsurge of interest in the special domain—often thought to be critical for intelligently controlled behavior—dealing with people's judgments about their own future knowledge and cognitive abilities (Blake, 1973; R. Brown & McNeill, 1966; Butterfield, Nelson, & Peck, 1988; Flavell, 1979; Gruneberg & Monks, 1974; Hart, 1965, 1967; Koriat & Lieblich, 1974; Lupker, Harbluk, & Patrick, 1991; Metcalfe, 1986; T. Nelson, Gerler, & Narens, 1984; T. Nelson & Narens, 1990; Reder, 1987, 1988; Schacter, 1983; Wellman, 1977, 1983). These are called *metacognitive judgments*. Our concern is in understanding the mechanisms and heuristics used in this special domain.

In a typical metamemory paradigm, people are asked the answers to a variety of memory questions. The manner in which they learned these answers initially may vary. On those questions on which they are unable to give a correct answer, they are requested to make a feeling-of-knowing judgment (FKJ), that is, they are asked to assess how likely

it is that they will be able to attain the answer in a subsequent test (that may or may not be different from the first test). People are willing to make these judgments. Under some circumstances, the assessments that they give show some accuracy: They know whether or not they will be able to remember an event later, even though they cannot remember it at the time of making the judgment (Blake, 1973; Hart, 1965, 1967; Metcalfe, 1986; T. Nelson, Leonesio, Shimamura, Landwehr, & Narens, 1982; Schacter, 1983). How people make these judgments is the specific concern investigated in the experiments that follow.

We initially consider two possibilities for the knowledge base of these judgments: first, familiarity with or knowledge about the cue, and second, partial information about the target. When we refer to target-based information, we mean correct information about the target itself. These mechanisms may not be mutually exclusive, of course. It is most unlikely that FKJs are based on factors entirely irrelevant to the task at hand. If they were, then manipulations that influence either cue familiarity or target retrievability would have no effect. Much data show that such variables do influence judgments (Blake, 1973; Koriat & Lieblich, 1977; Lupker et al., 1991; T. Nelson et al., 1982; Reder, 1987; Reder & Ritter, 1992; Schwartz & Metcalfe, 1992).

Probably the idea that FKJs are based on partial retrieval of the target itself is the most intuitive explanation for how these judgments are made. Even so, the empirical results favoring the idea are not overwhelming. Indeed, research from four independent laboratories points to the importance of the cues rather than the targets (Costermans, Lories, & Ansay, 1992; Glenberg, Sanocki, Epstein, & Morris, 1987; Metcalfe, 1993; Reder, 1987; Reder & Ritter, 1992; Schwartz & Metcalfe, 1992). Metcalfe (1993), in proposing that the basic memory system requires a monitoring and control mechanism to function normally, outlined a number of lines of evidence, mentioned below, suggest that people usually do not base their FKJs on partial retrieval of the target, but might instead use cue-related information.

First, in the standard paradigm for FKJs, subjects give estimates of the likelihood that they will later be able to

Janet Metcalfe, Bennett L. Schwartz, and Scott G. Joaquim,
Department of Psychology, Dartmouth College.

We gratefully acknowledge the support of National Institute of Mental Health Grant MH48066-03 to Janet Metcalfe. We thank Noga Gottesfeld for conducting a preliminary experiment in this series and for assisting with the programming and Jonathan Ellis for conducting Experiment 4. We also thank Asher Koriat for helpful comments and suggestions.

Correspondence concerning this article should be addressed to Janet Metcalfe, Department of Psychology, Dartmouth College, Hanover, New Hampshire 03755. Electronic mail may be sent to metcalfe@dartmouth.edu.

remember the answers to questions to which they cannot remember the answers at the time of the initial test. It seems reasonable to suppose that information other than the explicit target information, which they have just demonstrated they cannot retrieve, underlies the judgments. Second, there are differences in feeling of knowing among different error types, such that errors of omission are given lower FKJs than are errors of commission (Krinsky & Nelson, 1985). Presumably, if people were basing their judgments on what was retrieved, but the experimenter let them know that what had just been retrieved was wrong (as occurs in these experiments), their feeling-of-knowing ratings for those items should be especially low rather than especially high, as the data show. Metcalfe (1993) argued that the cues that evoke commission errors are probably more familiar than those that evoke nothing, because at least they were familiar enough to lead to some response. The cues for omission errors were so unfamiliar that they produced no response. This finding is consistent with the cue-familiarity hypothesis. Third, if FKJs were based only on retrieved information, then latencies to make these judgments would be slower than retrieval latencies. However, Reder (1987, 1988) found that FKJ latencies were faster than retrieval latencies. Fourth, Reder (1988) found that priming of the cues spuriously increased peoples' feelings of knowing without increasing their ability to answer the questions. Similarly, Glenberg et al. (1987) found a large positive correlation between peoples' domain knowledge, or the familiarity of the cues, and their FKJs, even in the absence of any accuracy of detailed predictions (see Costermans et al., 1992, and Maki & Serra, 1992, for related results). Reder and Ritter (1992) investigated a strategy-judgment paradigm in which subjects were given the choice of whether to retrieve or to calculate the results of arithmetic problems. The familiarity of the questions and the answers was varied. Their results pointed to the importance of cue familiarity. Fifth, Jameson, Narens, Goldfarb, and Nelson (1990) affected memory performance without changing the FKJs. They tachistoscopically (near the threshold of consciousness) presented subjects with some of the targets to general information questions. This resulted in an increase in recall, but had no discernible effect on FKJs. The fact that target retrievability may be altered without influencing FKJs suggests that the judgments are not based on target information. Sixth, as Kolers and Paley (1976) observed, people know what they do not know. It does not appear that they make these judgments by trying to retrieve what they do know because "don't know" judgments may be made very quickly. The fast latencies are consistent with judgments based on unfamiliar cues. Seventh, explicit "don't know" information results in impairment on "don't know" judgments. Glucksberg and McCloskey (1981) gave subjects statements indicating that certain information was not known (e.g., "It is not known whether Gabriel owns a saxophone"). If subjects were basing their judgments on what they retrieved, they should have been better at making "don't know" judgments as a result of this information. However, the experimental results showed that the explicit "don't know" information impaired performance, presumably because the explicit information made the cues seem more familiar.

In addition to these results, an initial result (Yaniv & Meyer, 1987) thought to favor the target retrievability hypothesis has recently been reexamined (Connor, Balota, & Neely, 1992) and appears instead to point toward the cue-familiarity hypothesis. Yaniv and Meyer (1987) gave subjects word definition problems such as "A statement that is seemingly contradictory or opposed to common sense and yet is perhaps true?" (paradox). Subjects were required to give TOTs and FKJs on those problems they could not immediately solve. These judgments were found to be correlated with quickened reaction time on a subsequent lexical-decision task on the targets. They also related to later recognition reaction time. Yaniv and Meyer interpreted these results as indicating that the questions had partially activated the representations of the targets, that these targets were more speedily processed later, and that the FKJs were based on the partial activation. This interpretation was challenged by Connor et al. (1992), who replicated Yaniv and Meyer's results, but who also found that the effect occurred even if the question and the FKJs were made 1 week after the tests of lexical decision and recognition. Speeded reaction time as a result of partial activation of the target from the test that had not yet occurred was not a viable explanation of the data. Connor et al. argued that the FKJs were made on the basis of familiarity with the domain of the cues. For domains that were highly familiar (and hence had high FKJs) subjects tended to respond quickly, but in domains about which the subjects knew little (and hence gave low FKJs) the times were accordingly slow.

The most compelling extant data indicating a target-based locus of metacognitive judgments revolves around the tip-of-the-tongue (TOT) feeling. TOTs refer to the subjective feeling that people sometimes have of being on the verge of recalling the target word. In contrast, FKJs are a voluntary estimate of the likelihood of recognizing the target word. The mere fact that people do experience TOTs (A. Brown, 1991; R. Brown & McNeill, 1966; Koriat & Lieblich, 1974) and that they are sometimes able to report the first letter or the number of syllables of the to-be-remembered word when they are in a TOT state suggests that partial information is available to them. If it is available, then it might be causally connected to the TOT states and to FKJs. A few experiments have produced data consistent with this idea. Blake (1973) showed that partial recall of three-letter trigrams was related to higher FKJs. Schacter and Worling (1985) demonstrated that subjects could recall better the affective valence of unrecalled items given high FKJs than those given lower FKJs. Although suggestive, both studies were correlational in nature and did not manipulate the availability of partial information. Indeed, Schacter (1981) reported a correlation between cue and target recognition, so these results could have been due to either factor. Because TOT responses appear to be of critical importance to the target retrievability hypothesis, we shall explore them in Experiments 3 and 4.

Some potential evidence for the target-retrievability hypothesis was also reported by Schwartz and Metcalfe (1992). In their experiments, cue familiarity was varied by a priming manipulation, following the lead of Reder (1987). Retrievability was manipulated by having subjects either read or

generate a rhyming word to the cue words (Slamecka & Graf, 1978) or read an unrelated word or by target priming. In three of the four experiments, the magnitudes of the FKJs were affected by the cue-familiarity manipulation but not by the target retrievability manipulation. In one experiment, though, priming either the cue or the target after the initial recall task affected the feeling of knowing. The target priming might have been thought to have affected partial target retrievability. Although Schwartz and Metcalfe argued that certain peculiarities of the experiment produced the target retrievability effects on FKJs, and although when they altered them in a subsequent experiment no effect of target priming on FKJs was found, it still seems prudent to explore further the idea that only cue familiarity underlies FKJs. The data, although suggestive, are not yet conclusive, and which factors may have the greatest impact on FKJs with different tasks are not yet known.

The present investigation had several purposes. The first was to contrast the cue-familiarity versus the target-retrievability hypothesis in a way different from that of either Reder (1987), Reder and Ritter (1992), or Schwartz and Metcalfe (1992). In the first two experiments presented here we capitalized on the fact that the standard conditions within the interference theory paradigm (e.g., A-B A-B, A-B A-B', A-B A-D, A-B C-D) allowed us to vary cue repetition (i.e., the number of repetitions of A) in a manner that is not linked to memorability of the target (B, in this case). (In the experiments that follow we shall switch the notation so that the target B is always second. Thus we will refer to the A-D A-B condition. This is done so that it is clear that the nature of the target itself is the same in all conditions and refers to the last pairing with the cue.) The second purpose of these experiments was to investigate the relation between tip-of-the-tongue feelings and feelings of knowing. Whether TOT feelings are related to FKJs and whether the same or different factors influence TOTs as FKJs is investigated in Experiments 3 and 4.

Experiment 1

In this experiment cue familiarity and target retrievability were both varied by taking advantage of the recall and repetitions relations in classic interference theory paradigms. Subjects were presented with a cued-recall list in which the cue A was paired with a target word B. They were told that they were to remember B such that when later given A they would be able to recall or recognize the item with which it was paired. Then they were given the second half of the list in which the relations depended on the first half. Sometimes the A items were given again and paired with the identical terms with which they were initially paired (A-B A-B). Sometimes the A term was paired with an item similar but not identical to the original B term (A-B' A-B). Sometimes the A term was paired with a new and unrelated word (A-D A-B). Finally, sometimes the A term was not repeated, but an entirely new cue-target pair was given in its stead (C-D A-B). The transfer relations on the second paired-associate list are well established (Martin, 1965; Osgood, 1949). Best transfer, as measured by recall of B, normally occurs in an

A-B A-B situation. The A-B' A-B situation results in the next best transfer. The C-D A-B list structure is traditionally considered to be the control condition in which no transfer occurs. Finally, an A-D A-B relation results in negative transfer.

The paradigm is especially interesting for the purposes of distinguishing the effects of memorability or target retrievability from cue familiarity on FKJs because the number of times the cue is repeated does not reflect memorability. Specifically, despite the fact that the cue is repeated twice in the A-D A-B condition, recall is expected to be the worst in that condition. In the C-D A-B condition, the cue is only presented once, but memorability is expected to be better in that condition than in the maximum interference condition, although certainly not as good as in the A-B A-B condition. So the rank ordering of the conditions if FKJs were based on target memorability should be A-B A-B > A-B' A-B > C-D A-B > A-D A-B. On the other hand, the rank ordering of conditions if cue repetition were critical in feeling of knowing judgments should be A-B A-B = A-B' A-B = A-D A-B > C-D A-B.

Method

Subjects. The participants in this experiment were 30 introductory psychology students at Dartmouth College who were given partial course credit in exchange for their participation. Subjects were tested individually in 1-hr sessions.

Materials. The words used in the lists were common two-syllable words that were selected at random from the Toronto Word Pool (Murdoch, 1968). In the A-B' A-B condition, the B and B' words were synonyms of one another. The closest possible synonyms were chosen (by scanning a dictionary). Examples were *lucky/fortunate* or *enough/sufficient*. Subjects were presented with a single list of cue-target pairs. The second half of the list included the 48 pairs that made up the critical pairs in each of the four conditions, such that there were 12 pairs in each condition. Materials were presented by Macintosh computers to subjects in a random order that maintained the first-half-second-half distinction.

Procedure. Subjects were instructed that they were to learn a list of pairs of items in such a way that when later presented with the left item (i.e., the cue) they could recall the right item (i.e., the target). If a particular cue was paired with more than one item, the subjects were instructed to recall the last item with which it was paired. The 96 item pairs (which included both repetitions within the four conditions) were given at a rate of 2 s per pair with a 2-s pause between presentations. Following presentation, subjects were given the cue items and asked to type in the targets. The computer scored, on-line, whether the answer was correct or incorrect.

For the incorrect items, subjects were asked to make a feeling-of-knowing judgment on a scale from 1 to 100, on which 100 meant that the subjects were sure they would be able to recognize the correct target when presented with a forced-choice recognition test and 1 meant that they had no idea what the correct answer would be and would be unable to recognize it. In this experiment, though not in those that follow, the program disallowed the same numerical judgment. We did this here so that there would be no ties in the rankings of judgments. We did not specifically allude to a guessing probability (though occasionally a subject would ask about it and the experimenter would answer the question).

Subjects were then given an eight-alternative forced-choice recognition test. The target was one of the alternatives and was placed in a random position with respect to the other options. The other

choices were words randomly chosen from the Toronto Word Pool with the constraint that they did not occur elsewhere in the experiment. We computed mean FKJs, gammas, recall, and cued recognition proportions for all conditions on each subject. The summary data for each measure on the basis of each subject comprised the data entered into the analyses of variance reported below.

Results

In the Results section for this and subsequent experiments, the criterion for considering an effect or interaction to be significant was set at a p value less than or equal to .05, as adjusted by Greenhouse-Geisser epsilon. P values will, therefore, not be stated (unless we are discussing a trend, i.e., $.05 < p < .10$) that did not quite reach our designated cutoff for significance. The Greenhouse-Geisser epsilon measures the extent to which the correlation between observations violates the assumption of sphericity, which is required for the univariate measures hypothesis test (Stevens, 1986). The memory results will be presented first; the metamemory results will follow.

Recall. Table 1 shows the level of cued recall as a function of experimental condition. The results are consistent with the findings from interference theory and with our expectations. The A-B A-B condition showed the highest level of recall ($M = .39$), followed by the A-B' A-B condition (.33), the C-D A-B condition (.19), and finally the A-D A-B condition (.17), $F(3, 87) = 19.53$, $MS_e = .02$. Newman-Keuls post hoc tests (Winer, 1971, p. 528) revealed that the A-B A-B and A-B' A-B conditions were significantly different from the C-D A-B and A-D A-B conditions.

Recognition. The recognition hit rates were: .72 for the A-B A-B condition, .56 for the A-B' A-B, .59 for A-D A-B, and .48 for the C-D A-B condition, $F(3, 87) = 13.00$, $MS_e = .02$. Newman-Keuls post hoc tests indicated that the C-D A-B condition was lower than the other three conditions combined, and the A-B A-B condition was higher than each of the other three conditions. These results are also shown in Table 1.

FKJs. As can be seen in Figure 1, the average magnitude of the FKJs did not reflect the memorability of the conditions, as given by either the initial recall test or the final recognition test, but rather mirrored the number of times the cue was repeated. In those conditions that repeated the cue FKJs were high, whereas in the condition in which the cue was presented only once the average feeling-of-knowing magnitude was low, $F(3, 87) = 12.64$, $MS_e = 83.13$. Newman-Keuls post

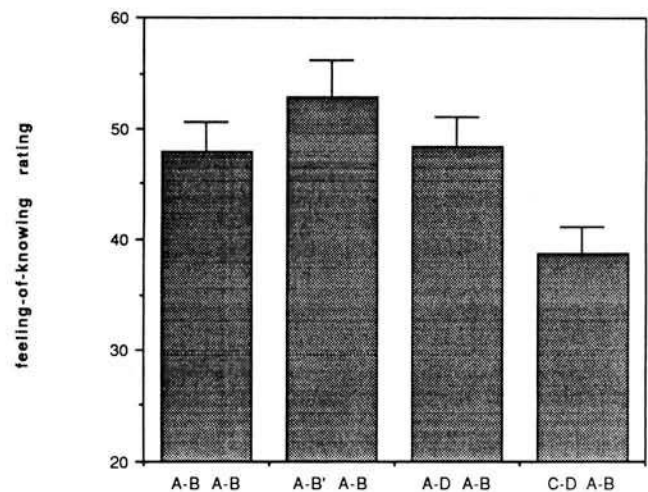


Figure 1. Magnitude of feeling-of-knowing judgments, on a scale from 1–100, in Experiment 1.

hoc tests showed that only the C-D A-B condition differed from the others.

Correlation between feeling of knowing and recognition. Gamma correlations (see Nelson, 1984) relating the FKJs to recognition accuracy are not a major concern of this article, but because these results may be of archival value we report the correlations for this and subsequent experiments in the Appendix.

Discussion

The magnitude of the FKJs reflected the number of times the cue had been presented. Thus, the results of this experiment provided support for the idea that manipulated cue familiarity is linked to feeling-of-knowing estimations. Our measures of memorability—initial recall and later recognition—showed somewhat different patterns, especially in the A-D A-B condition, which was lowest in recall but in the middle of the other conditions in recognition. It is not clear which of these two measures is more appropriate as an indicator of memorability of the target (though, of course, both of these were different from the pattern of FKJs). The recall measure, though showing a memorability pattern consistent with past literature, suffers from the fact that it includes the recalled items—items that were excluded when FKJs were made. To allow that the initial recall data accurately reflect the memorability of the unrecalled targets we have to assume that the unrecalled items in each condition were affected by the manipulation in the same way as were the recalled items. To make this inference, one would like corroborating evidence from the second test. However, the recognition results were somewhat different from those of the initial recall. Even so, the pattern of metacognitive ratings did not reflect *either* memorability measure. Instead, the FKJs reflected the number of times the cue was repeated, consistent with the cue-familiarity hypothesis.

We thought that one problem with the recognition measure as an unbiased measure of memory was the nature of the

Table 1
Recall and Recognition Performance Means
in Experiment 1

Condition	Recall ^a	Recognition ^b
A-B A-B	.39	.72
A-B' A-B	.33	.56
A-D A-B	.17	.59
C-D A-B	.19	.48

^a Recall refers to percentage correct in the initial recall test. ^b Recognition refers to percentage correct in eight-alternative forced-choice recognition of the unrecalled items.

lures. This has been shown in a number of other experiments (e.g., Blake, 1973; McCloskey & Zaragoza, 1985) to be an important factor in determining the outcome of the test. For example, in Experiment 1 correct recognition could occur solely on the basis of the familiarity of the target, rather than because subjects knew about the relation between the cue and the target. In addition, the items that might cause interference were not included among the alternatives and therefore could not interfere. This exclusion may have resulted in the unexpectedly high level of recognition in the A-D A-B condition. We designed the second experiment to attempt replication of the results of Experiment 1 on FKJs and to investigate this possibility concerning recognition.

Experiment 2

The second experiment was similar to the first, except that we altered the nature of the lures used at time of recognition to make the recognition situation more like the original recall situation and so that the recognition test would more reasonably reflect what was retrievable and confusable in memory. Specifically, the forced-choice recognition alternatives included the items expected to cause interference with the target. We expected that, because these alternatives presumably could not be eliminated easily in recall, this choice of lures would produce a recognition pattern different from that found in the first experiment and more like that found in recall. Our hypothesis, though, was that the FKJs would still be directly related to the number of repetitions of the cues (as in Experiment 1) and not to target retrievability as measured by either recall or recognition.

Method

The subjects were 24 Dartmouth College undergraduates taking an introductory psychology course. Subjects participated in exchange for extra credit in their course. The method was the same as that of Experiment 1 with several exceptions. The most important change was in the structure of the lures in the final recognition phase. In this experiment the recognition test alternatives were the target B item; a close associate of the target (which in the A-B' A-B condition was the B' item but in the other conditions was an un-presented associate); an unrelated word (which in the A-D A-B condition was the D word and in the other conditions was an extralist lure); a close associate of the unrelated word (so that the subject could not guess the target by means of the uneven construction of the test alternatives); an intralist word, but one that was not paired with the cue A (i.e., the so-called D word in the C-D A-B condition); and a close associate of the intralist word. Each subject participated in four sessions with new words in each session. The B' items in this experiment were associates, rather than synonyms, as in the first experiment. The associates were selected from the norms of D. Nelson, McEvoy, and Schreiber (1989). Each associate chosen was neither a synonym nor an antonym of the first word, but each still resulted in higher than 3% production rates in the free association norms. The order of tasks was study, initial recall, FKJ task, and final recognition.

Table 2
*Recall and Recognition Performance Means
in Experiment 2*

Condition	Recall ^a	Recognition ^b
A-B A-B	.56	.71
A-B' A-B	.37	.42
A-D A-B	.23	.40
C-D A-B	.24	.54

^a Recall refers to percentage correct in the initial recall test. ^b Recognition refers to percentage correct in six-alternative forced-choice recognition of the unrecalled items.

Results

There were no interactions between sessions and any of the manipulated variables on any of the dependent variables, therefore the data were collapsed across sessions. The memory tests are presented first, followed by the metamemory results.

Initial recall. Recall from this experiment is shown in Table 2. As had been the case in the first experiment, there were differences between conditions in recall, $F(3, 69) = 146.54$, $MS_e = .00$. Newman-Keuls post hoc tests revealed that the A-B A-B condition was best, followed by the A-B' A-B condition. In this experiment, recall in the A-D A-B condition and in the C-D A-B condition were not significantly different from one another.

Recognition. The recognition test produced a different pattern from that found in Experiment 1, as is shown in Table 2. There was an overall effect of the conditions on recognition, $F(3, 69) = 50.93$, $MS_e = .01$. Newman-Keuls post hoc tests showed that the targets were recognized best in the A-B A-B condition and next best in the C-D A-B condition. In both the A-B' A-B condition and the A-D A-B condition recognition was low, and these two were at the same level. The reason for this pattern of recognition is undoubtedly attributable to the nature of the lures. In the two conditions that showed high recognition, there was no alternative in the test set that had been paired with the cue. Recognition in the A-B A-B condition had an advantage because the pair had been repeated. In the other two conditions, however, the distractor that had initially been presented with the cue was one of the alternatives present in the test. It is reasonable to suppose that interference from this alternative was responsible for the low level of recognition.

FKJs. As had been the case in Experiment 1, the FKJs were related to the number of repetitions of the cues, rather than to the level of recall or recognition, $F(3, 69) = 20.99$, $MS_e = .00$. Newman-Keuls post hoc analysis showed that the A-B A-B condition (.42), the A-B' A-B condition (.46), and the A-D A-B condition (.48), conditions in which the cue was presented twice, all showed higher mean FKJs than did the C-D A-B condition (.37), in which the cue was presented only once. This pattern is shown in Figure 2.

Discussion

This experiment, like the first, showed that the magnitude of the FKJs mirrored the number of repetitions of the cue and

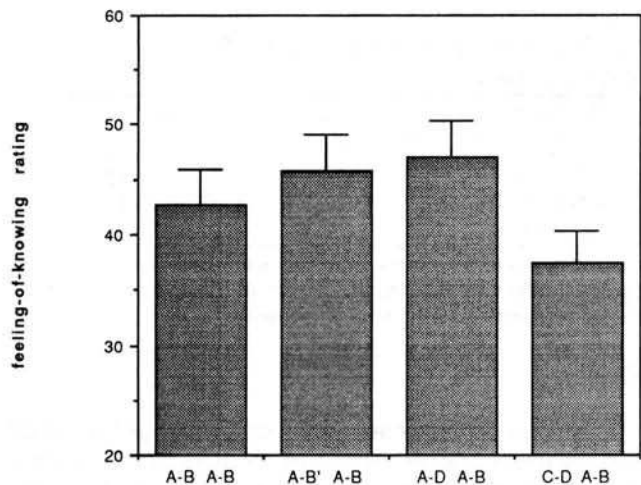


Figure 2. Magnitude of feeling-of-knowing judgments, on a scale from 1–100, in Experiment 2.

not the memorability of the targets, whether that memorability was measured by the initial recall or by the later recognition test. The initial recall and the recognition test showed more similar patterns in this experiment than they had in Experiment 1. They also more closely conformed to the classic transfer patterns expected by interference theory. In particular, the A-B A-B condition was best, the C-D A-B condition was in the middle, and the A-D A-B condition was worst. There was a discrepancy in the A-B' A-B condition such that in recall this condition was high, whereas in recognition it was low. This is an unusual condition because performance in it might depend on the exact relations between the items. The inclusion of the highly associated lures no doubt contributed to the low performance in recognition.

The results of Experiment 2 were consistent with those of Experiment 1. The three conditions showing high feelings of knowing were those in which the cues were repeated, regardless of the subsequent memorability of the targets in those conditions. The condition in which the feeling-of-knowing magnitude was low was the condition in which the cue was presented only once. These results are consistent with the cue-familiarity hypothesis.

Experiment 3

The tip of the tongue state is often thought to be the strongest evidence for the hypothesis that people have partial information about the target, even when they cannot retrieve the target, and, by implication, that partial target information might be responsible for feeling-of-knowing judgments. In Experiments 3 and 4 we are concerned, then, not only with the mechanisms underlying feeling-of-knowing judgments but also with those concerned with tip-of-the-tongue states.

Although it might seem intuitively obvious that the existence of the TOT phenomenon implies partial target information, nevertheless, within the literature there are two

explanations of the phenomenon, only one of which implicates partial target information. The first traditional hypothesis, which will be described in more detail shortly, is the *ugly stepsister* (or blocking) hypothesis. The second is the *partial information hypotheses*. A third hypothesis, and a possibility of concern here because it relates TOT responses to FKJs, is the *cue familiarity hypothesis*. This third hypothesis has received little attention in the TOT literature except in one study by Koriat and Lieblich (1977), which showed that different memory pointers or questions differentially evoked TOT reports.

The ugly stepsister hypothesis states that the TOT feeling is a manifestation of the blocking of the correct response by incorrect intruders, rather than of partial access to the correct response itself (A. Brown, 1991; Jones, 1989). A. Brown (1991, p. 215) wrote "the blocking perspective suggests that the TOT represents a memory search that has become sidetracked in the wrong memory location." A variant of this hypothesis was named the ugly stepsister hypothesis by Reason and Lucas (1984). According to this hypothesis, the subject tries to retrieve the correct response (Cinderella), but the incorrect alternatives (the ugly stepsisters) appear and block and interfere with the attainment of this preferred response. This explanation suggests that the information subjects can retrieve when they are in a TOT state should relate to intruders rather than to the target. In keeping with this hypothesis, Jones (1989) found that presentation of words phonologically related to the targets increased the reported number of TOT states.

The partial correct target information hypothesis proposes that the TOT state results because people have retrieved some aspects or features of the target item. This information, though, is insufficient to allow the entire item to be reconstructed at the time of the memory query. The successful retrieval of these target fragments is thought to underlie the TOT feelings and is said also to underlie subjects' above-chance knowledge of the first letter and number of syllables in the target (see A. Brown, 1991, for a review of the literature).

The TOT feeling seems akin to feelings of knowing. Certainly, one might consider that if a subject were in a TOT state, he or she would be likely to give the item in question a high feeling-of-knowing rating. Insofar as TOT states are based on partial retrieved information, then it would seem reasonable to suggest that at least some of these FKJs might, by extension, also be based on retrieved target information. Alternatively, we might find that there is a relation between TOT states and FKJs but that both are based on cue familiarity.

In the third experiment we investigated three hypotheses of the locus of tip-of-the-tongue feelings, using the three major interference theory conditions—A-B A-B, A-D A-B, and C-D A-B—as in the preceding experiments. First, the TOT feeling might be based on partial target information. By this partial target retrieval hypothesis, it would be expected that TOT reports would be most prevalent in the A-B A-B condition, next most prevalent in the C-D A-B condition, and least prevalent in the A-D A-B condition, that is, they would follow the memorability pattern.

Second, the TOT state might be the same as or closely related to the feeling-of-knowing state and so might be influenced by the same factors that influence FKJs. Previous experiments showed that FKJs are sensitive to the familiarity of the cues. Thus, the second hypothesis is that TOT states, like FKJs, will be based on cue familiarity.

The third possibility is the ugly stepsister hypothesis. In only the A-D A-B condition was an alternative presented that would be expected to block the target. Thus, only the A-D A-B condition should show a high frequency of TOT feelings. In summary, then, the frequency of TOT feelings by condition predicted by the three hypotheses are (a) target retrievability: A-B A-B > C-D A-B > A-D A-B; (b) cue familiarity: A-B A-B = A-D A-B > C-D A-B; and (c) ugly stepsister: A-D A-B > C-D A-B = A-B A-B.

Method

Subjects. The subjects were 30 undergraduates recruited from an introductory psychology course at Dartmouth College. They received extra credit in the course for participation in the experiment. Subjects were tested individually for approximately 1 hr each.

Design. The design was the same as that used in Experiment 1, with the following exceptions. Most important, TOT judgments and first-letter identification were included. In addition, we eliminated the A-B' A-B condition.

The alternatives consisted of the correct target and seven new distractors. All words were randomly chosen from the Toronto Word Pool (Murdock, 1968), without replacement. Each subject received a different randomly selected list.

Procedure. As in the preceding experiment, subjects first studied the list of paired associates. They were instructed that during the recall phase if they could not recall the target they would be asked if they were experiencing a tip-of-the-tongue state, which was described as being a state in which one felt "on the verge of recalling the target word." Subjects were also told to guess the first letter of the target word. Thus the TOT judgments and the first-letter identification were done on a word-by-word basis linked to the initial cued-recall task. After having gone through all cues for recall, the cues were rerandomized and subjects were requested to make FKJs, just as in Experiments 1 and 2. They were then given an eight-alternative forced-choice recognition test. The order of tasks, then, was (a) initial study; (b) cued-recall, TOT judgment, and first-letter identification, done on a word-by-word basis; (c) the FKJs task, done as a block; and (d) the final recognition task, also done as a block.

Results

As before, we shall first present the memory results. Then the metamemory data, here comprising both FKJs and TOT data, will be reported. The first-letter identifications were considered to be memory data (though presumably partial, rather than complete, retrieval).

Recall. There was a main effect of encoding condition for recall, $F(2, 58) = 47.75$, $MS_e = .01$, shown in Table 3. Newman-Keuls post hoc tests revealed that items from the A-B A-B condition were most likely to be recalled, followed by the C-D A-B condition and the A-D A-B condition.

Recognition. Table 3 shows the recognition results, $F(2, 58) = 31.56$, $MS_e = .01$. Newman-Keuls post hoc tests

showed that as in recall, the A-B A-B condition was best, followed by the A-D A-B and the C-D A-B conditions. In recognition, these latter two conditions did not differ significantly from one another.

First-letter identification. The proportion of correct first letters was .16 in the A-B A-B condition, .09 in the A-D A-B condition, and .08 in the C-D A-B condition, $F(2, 58) = 5.00$, $MS_e = .01$. The pattern follows that of recall and recognition. Correct first-letter guesses, then, mirrored target retrievability in this experiment.

FKJs. As is shown in the left panel of Figure 3, there was a main effect of encoding condition on the FKJs, $F(2, 58) = 9.78$, $MS_e = 72.06$. As had been the case in the two preceding experiments, FKJs mirrored the number of repetitions of the cue A. Newman-Keuls post hoc tests showed that the A-B A-B condition and the A-D A-B condition had higher FKJs than did the C-D A-B condition.

TOT. The TOT data were like the FKJ data. The frequency of reported TOT feelings was high in the repeated-cue conditions and low in the singly presented cue condition, $F(2, 58) = 9.80$, $MS_e = .01$. These results are shown in the right-hand panel of Figure 3.

We conditionalized the data as a function of whether subjects were or were not in a TOT state. Because not all subjects had conditionalized data in all conditions, the degrees of freedom vary in some of the following analyses. When in TOT states, subjects were more accurate in identifying the correct first letter and in recognizing the target word, $F(1, 42) = 19.01$, $MS_e = .05$, and $F(1, 44) = 16.16$, $MS_e = .05$, respectively. Higher feelings-of-knowing ratings were given for items in which subjects reported being in TOT states than for those in which they denied a TOT feeling, $F(1, 44) = 70.38$, $MS_e = 443.60$. TOT states showed significantly higher gamma correlations between FKJs and recognition than did non-TOT states, $F(1, 27) = 12.68$, $MS_e = .14$.

Discussion

This experiment, like the first two, showed that FKJs depended on the number of repetitions of the cue. The TOT data also depended on cue familiarity and not target memorability. The pattern of these data support the cue familiarity hypothesis. They are not consistent with either the target retrievability hypothesis or the ugly stepsister hypothesis.

Experiment 4.

The fourth experiment was like the third experiment except that the final test was cued recall rather than cued recognition. There were two reasons for using the recall test. First, there was some ambiguity about memorability across conditions using recognition. As we and others have shown, the nature of lures in the recognition test has a pronounced effect on the inferences one would make concerning goodness of memory across conditions. We were therefore concerned whether the pattern of memorability in a second recall test would be the same as in a first test.

Second, we wished to explore a conjecture by Koriat (1992) concerning FKJs that came to light while we were

Table 3
*Recall and Recognition Performance Means
 in Experiment 3*

Condition	Recall ^a	Recognition ^b
A-B A-B	.40	.69
A-D A-B	.13	.51
C-D A-B	.19	.47

^a Recall refers to percentage correct in the initial recall test. ^b Recognition refers to percentage correct in eight-alternative forced-choice recognition of the unrecalled items.

conducting these experiments. We shall call this conjecture the *omnibus retrieval hypothesis*, which states that FKJs are based on the total amount of information that is retrieved, regardless of whether the information is correct or incorrect. The forced-choice recognition procedure did not allow assessment of the total amount retrieved. Because in the recognition test the subjects were forced to choose exactly one alternative as the response to every cue, we forced the amount of information retrieved to look the same in all conditions. The cued-recall task, however, allows the subject to freely produce correct information, incorrect information, or nothing, and thus allows us to examine the total information retrieved in each treatment combination.

A third reason for conducting this experiment was unrelated to the nature of the final test. We wished to know if the TOT and first-letter results reported in Experiment 3 were replicable. Our hypothesis, of course, was that we would find the same pattern of metamemory results in Experiment 4 as in Experiment 3.

Method

The method was the same as that of Experiment 3 with the important change being that the second memory test was cued recall rather than recognition. We also included a primacy and a recency buffer of pairs of items, each consisting of 5 pairs, and an exper-

imenter, rather than the subjects, typed the responses into the computer. Subjects were 24 Dartmouth College students who received extra credit in an introductory psychology course in exchange for their participation. They were instructed that the FKJs were to relate to a later recall test (rather than to a recognition test).

Results

Memory results. Table 4 shows that there was an effect of condition on initial recall, such that the A-B A-B condition was better than either the A-D A-B or the C-D A-B condition, which did not differ from one another, $F(2, 46) = 48.04$, $MS_e = .01$. Table 4 also shows performance on the second recall test. Although the overall level of recall of the previously unrecalled items in the three conditions was lower than in the first test, the pattern was the same as had been found on the first test, $F(2, 46) = 15.90$, $MS_e = .01$. Thus, the measure of memorability seemed quite stable in this experiment. The correct first-letter responses (A-B A-B = .11, A-D A-B = .13, C-D A-B = .11) showed no significant differences among conditions in this experiment ($F < 1$). Thus, although the two recall tests show a consistent pattern of memorability, the first-letter judgments in this experiment reflected neither the basic memorability of the conditions nor the metamemory judgments.

Metamemory results. As had been the case in Experiment 3, the FKJs mirrored the number of repetitions of the cue rather than the memorability of the target. The A-B A-B condition and the A-D A-B conditions were high and the same, whereas the C-D A-B condition showed lower ratings, $F(2, 46) = 5.28$, $MS_e = 57.47$. These data are shown in the left panel of Figure 4. Shown in the right panel are the TOT data. Once again, the TOT data showed a relation to repetitions of the cues rather than to the memorability of the targets, $F(2, 46) = 7.06$, $MS_e = .02$.

Additional analyses relating to the omnibus retrieval hypothesis. The omnibus retrieval hypothesis states that FKJs and TOT judgments are based on all retrieved information

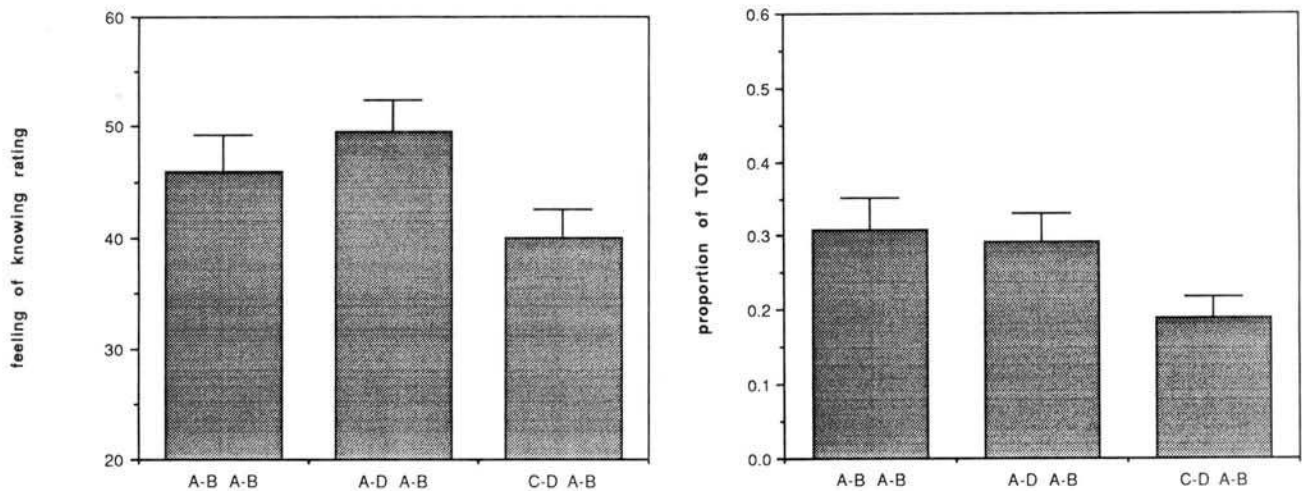


Figure 3. The left panel shows magnitude of feeling-of-knowing judgments, and the right panel shows proportion of tip-of-the-tongue (TOT) states, in Experiment 3.

Table 4
Recall and Second Recall Means in Experiment 4

Condition	Recall ^a	Second recall ^b
A-B A-B	.50	.17
A-D A-B	.19	.05
C-D A-B	.22	.06

^a Recall refers to percentage correct in the initial recall test.
^b Second recall refers to cued recall for initially unrecalled items.

and not just correct information. To assess this hypothesis, we reanalyzed the second recall data by condition including not only correct recall but also intrusions. We expected to find that these total recall measures would reveal the same pattern as FKJs and TOT judgments. The data are shown in Table 5. The results show that when all retrieved items, whether correct or incorrect, are included, the C-D A-B condition showed less total retrieval than did either of the A-B A-B or the A-D A-B conditions, which were not different from one another. This pattern is the same as that which was found for the FKJs and the TOT data. Thus, in this experiment, using this particular measure of retrievability, the omnibus retrieval hypothesis was correlated with the cue familiarity hypothesis, and both made the same predictions—predictions that were supported by the data.

Discussion

This experiment again provided support for the cue familiarity hypothesis for both FKJs and TOT judgments and contradicted the target retrieval hypothesis. The second recall data from this experiment also showed that correct plus incorrect retrieval was directly related to the number of cue repetitions. This correlation need not occur in all cases, but did occur in this experiment, and so these results can be explained not only by the cue familiarity hypothesis but also by the omnibus retrieval hypothesis (Koriat, 1992). The two

hypotheses appear to be almost perfectly related to one another in this experiment.

Conclusion

Our concern in this article was in determining how people make metamemory judgments, and, in particular, how they make feeling-of-knowing and tip-of-the-tongue judgments. During the course of the four experiments, we tested four hypotheses, namely the target retrievability hypothesis, the cue familiarity hypothesis, the ugly stepsister hypothesis, and the omnibus retrieval hypothesis. All of the data from the four experiments was consistent with the cue familiarity hypothesis. In each of the four experiments, the familiarity of the cue was varied in a manner that was dissociated from the retrievability of the target. In each case, the familiarity of the cue mapped onto the pattern of metacognitive judgments, whereas the retrievability of the correct target did not. This finding indicates that the judgments are not normally based on target retrievability. The judgments also did not appear to be especially low when no blockers were present, that is, when the target was unambiguously remembered. Indeed the judgments were typically high in this condition. We therefore conclude that the ugly stepsister or blocking hypothesis does not illuminate our data. Finally, the omnibus retrieval hypothesis proposes that total retrieval, and not just correct retrieval, should relate to the magnitude of the metacognitive judgments. Like the cue familiarity hypothesis, we did find some evidence in support of this hypothesis. We do not yet have enough data on the omnibus retrieval hypothesis, however, to conclude anything about its status except that further research is needed. In all of the experiments reported here, the number of repetitions of the cue was diagnostic of the metacognitive judgments—both FKJs and TOT judgments. In the one experiment that allowed us to investigate the possibility, repetition of the cue was also correlated with the total amount of information retrieved by that cue, though other

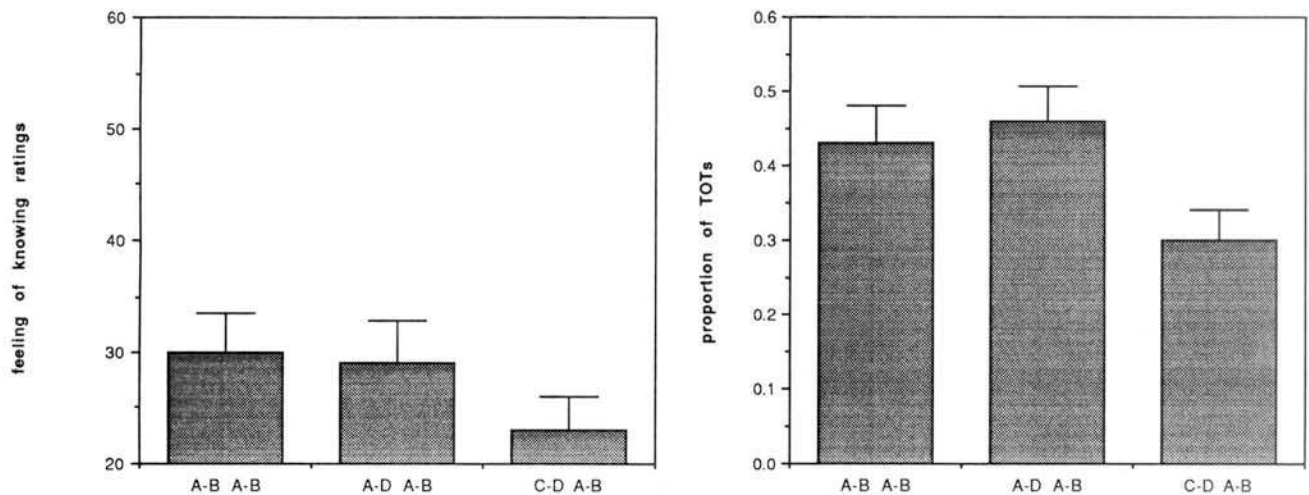


Figure 4. The left panel shows magnitude of feeling-of-knowing judgments, and the right panel shows proportion of tip-of-the-tongue (TOT) states, in Experiment 4.

Table 5
Omnibus Retrieval: Second Recall Means
in Experiment 4

Condition	% correct	% old ^a	% intrusions ^b	Omnibus ^c
A-B A-B	.17	—	.35	.52
A-D A-B	.05	.10	.34	.49
C-D A-B	.06	—	.32	.38

Note. Dashes = data not applicable in these conditions.

^a Percent old refers to percentage of times the D items were recalled.

^b Percent intrusions refers to the percentage of times a commission error occurred that was not the D item.

^c Omnibus refers to total amount retrieved.

data collected in our lab (Schwartz & Metcalfe, 1992) fail to show a similar pattern favoring the omnibus retrieval hypothesis. The simplest explanation of our data, to date, is that the metamemory judgments are based on cue familiarity. Because this means of assessment is indirect, we shall call it the *cue familiarity heuristic*.

In making judgments related to the external world rather than to their internal states and abilities, people use other heuristics or rules of thumb. These heuristics approximate the uncertain quantity indirectly, rather than measuring the quantity itself. Because they do not measure the quantity directly, such heuristics may result in biases and errors, the specification of which has generated much excitement. We argue here that feeling-of-knowing judgments are also made by using an heuristic—specifically, by assessing the familiarity of the cue (rather than by evaluating the quantity or to-be-remembered event directly, i.e., by assessing the target itself). Within the internally focused domain of metamemory judgments we find that sometimes the judgments are accurate, but frequently our predictive ability is disappointing, as might well be expected given that we base these judgments on a heuristic rather than on some direct assessment of the quantity or target itself. Use of the cue familiarity heuristic may result in correct judgments so long as the familiarity of the cue is correlated with correct target retrieval (as it often is—familiar cues tend to be paired with memorable targets). However, this cue familiarity need not always correlate with target recallability, and when it does not, its use may result in specific and analyzable mistakes.

References

- 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.
- Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin*, 109, 204–233.
- Brown, R., & McNeill, D. (1966). The “tip of the tongue” phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 5, 325–337.
- Butterfield, E. C., Nelson, T. O., & Peck, V. (1988). Developmental aspects of the feeling of knowing. *Developmental Psychology*, 24, 654–663.
- Connor, L. T., Balota, D. A., & Neely, J. H. (1992). On the relation between feeling of knowing and lexical decision: Persistent sub-threshold activation or topic familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 544–554.
- Costermans, J., Lories, G., & Ansary, 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.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-development inquiry. *American Psychologist*, 34, 906–911.
- Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, 116, 119–136.
- Glucksberg, S., & McCloskey, M. (1981). Decisions about ignorance: Knowing that you don't know. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 311–325.
- Gruneberg, M. M., & Monks, J. (1974). Feeling of knowing and cued recall. *Acta Psychologica*, 38, 257–265.
- Hart, J. T. (1965). Memory and the feeling of knowing experience. *Journal of Educational Psychology*, 56, 208–216.
- Hart, J. T. (1967). Memory and the memory-monitoring process. *Journal of Verbal Learning and Verbal Behavior*, 6, 685–691.
- Jameson, K. A., Narens, L., Goldfarb, K., & Nelson, T. O. (1990). The influence of near-threshold priming on metamemory and recall. *Acta Psychologica*, 73, 55–68.
- Jones, G. V. (1989). Back to Woodworth: Role of interlopers in the tip-of-the-tongue phenomenon. *Memory & Cognition*, 17, 69–76.
- Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, England: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251.
- Kolers, P. A., & Palef, S. R. (1976). Knowing not. *Memory & Cognition*, 4, 553–558.
- Koriat, A. (1992). *How do we know that we know? The accessibility account of the feeling of knowing*. Manuscript submitted for publication.
- 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.
- Lupker, S. J., Harbluk, J. L., & Patrick, A. S. (1991). Memory for things forgotten. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 897–907.
- Maki, R. H., & Serra, M. (1992). The basis of test predictions for text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 116–126.
- Martin, E. (1965). Transfer or verbal paired associates. *Psychological Review*, 72, 327–343.
- McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General*, 114, 1–16.
- Metcalfe, J. (1986). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 288–294.
- Metcalfe, J. (1993). Novelty monitoring, metacognition, and control in a composite holographic associative recall model: Implications for Korsakoff amnesia. *Psychological Review*, 100, 3–22.
- Murdock, B. B. (1968). Modality effects in short-term memory: Storage or retrieval? *Journal of Experimental Psychology*, 77, 79–86.

- Nelson, D. L., McEvoy, C. L., & Schreiber, T. (1989). *The University of South Florida word association, rhyme and word fragment norms*. Unpublished manuscript.
- Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*, 109–133.
- Nelson, T. O., Gerler, D., & Narens, L. (1984). Accuracy of feeling of knowing judgments for predicting perceptual identification and relearning. *Journal of Experimental Psychology: General*, *113*, 282–300.
- Nelson, T. O., Leonesio, R. J., Shimamura, A. P., Landwehr, R. S., & Narens, L. (1982). Overlearning and the feeling of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 279–288.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–141). San Diego, CA: Academic Press.
- Osgood, C. E. (1949). The similarity paradox in human learning: A resolution. *Psychological Review*, *56*, 132–143.
- Reason, J. T., & 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). Selection strategies in question answering. *Cognitive Psychology*, *19*, 90–138.
- Reder, L. M. (1988). Strategic control of retrieval strategies. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 227–259). San Diego, CA: Academic Press.
- Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 435–452.
- Schacter, D. L. (1981). *Feeling of knowing and the expression of knowledge from episodic memory*. Unpublished doctoral dissertation, University of Toronto.
- Schacter, D. L. (1983). Feeling of knowing in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 39–54.
- Schacter, D. L., & Worling, J. R. (1985). Attribute information and the feeling of knowing. *Canadian Journal of Psychology*, *39*, 467–475.
- 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.
- Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 592–604.
- Stevens, J. (1986). *Applied multivariate statistics for the social sciences*. Hillsdale, NJ: Erlbaum.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *10*, 438–464.
- Wellman, H. M. (1977). Tip of the tongue and feeling of knowing experiences: A developmental study of memory monitoring. *Child Development*, *48*, 13–21.
- Wellman, H. M. (1983). Metamemory revisited. In T. H. Chi (Ed.), *Trends in memory development research* (pp. 31–51). Basel, Switzerland: Karger.
- Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill.
- 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.

Appendix

Correlation Between Feeling-of-Knowing Judgments and Recognition

Condition	Gamma scores ^a			
	Exp. 1	Exp. 2	Exp. 3	Exp. 4
A-B A-B	.03	.33*	.25*	.47*
A-B' A-B	.36*	.09	—	—
A-D A-B	.08	.04	.20*	.26*
C-D A-B	.21	.25*	.14	.44*

Note. Exp. = Experiment. Dashes = data not applicable in these conditions.

^a Gamma refers to the Goodman-Kruskal gamma correlation between the feeling-of-knowing judgments and subsequent recognition.

* Significantly different from zero at the $p < .05$ level.

Received June 30, 1992
 Revision received October 13, 1992
 Accepted October 20, 1992 ■