What Determines Initial Feeling of Knowing? Familiarity With Question Terms, Not With the Answer

Lynne M. Reder and Frank E. Ritter
Carnegie Mellon University

How do people know whether they have an answer to a question before they actually find it in their memory? We conducted 2 experiments exploring this question, in which Ss were trained on relatively novel 2-digit × 2-digit arithmetic problems (e.g., 23 × 27). Before answering each problem, Ss made a quick *feeling of knowing* judgment as to whether they could directly retrieve the answer from memory or had to compute it. Knowing the answer initially appeared to be linearly related to having a feeling of knowing the answer; however, when the frequency of exposure to complete problems and the frequency of exposure to parts of the problems were separately varied, feeling of knowing was better predicted by the frequency of presentation of the problem parts, not by knowledge of the answer. This suggests that the processes involved in knowing the answer are different from those involved in having a feeling of knowing. Specifically, an early feeling of knowing is not just based on an early read of the answer.

How do people go about answering questions? It is not necessarily as straightforward as it might seem. For example, one must infer the frame of reference of the questioner and adjust the degree of specificity of the answer to that frame of reference. Norman (1973), for example, noted that an acceptable answer to the question "Where is the Empire State Building?" depends on whether the questioner is asking it while in Europe, outside of New York City, or in the city itself. In addition to determining the questioner's frame of reference, there are other issues the answerer must consider in order to choose the optimal strategy for generating the answer. The answerer must decide whether he or she already knows the answer to the question, whether it can be inferred, whether search is needed to answer it, or whether there is no possible answer (e.g., "What was Shakespeare's phone number?").

This article is concerned with the issue of metaknowledge, that is, how people determine what they know about a question before they actually answer it. In particular, is it necessary to search memory in order to ascertain whether an answer is likely to be known? Although it might seem reasonable that people first search their memory for an answer before trying some other question-answering strategy, such as figuring out the answer, there is evidence that this is not the case (e.g.,

Reder, 1982, 1987). Question-answering strategy choices appear to be biased by variables of the questioning situation and by the question itself (Reder, 1987, 1988). What is not known, however, is whether people can use a very brief, initial evaluation of a question to guide strategy choice.

One possibility is that initial evaluation involves some sort of feeling of knowing process. Earlier researchers (e.g., Hart, 1965, 1967) defined feeling of knowing as only the state of believing that currently unrecallable information will be available (in some form) later because the knowledge is in memory. The tip-of-the-tongue phenomenon (Brown & McNeill, 1966) is a related state, in which the strength of feeling of knowing is very strong and the person believes that the information should be recallable but somehow is blocked momentarily. What both these feelings represent, essentially, is a decision that the answer is in memory and continued efforts at retrieval are worthwhile (Gruneberg, Monks, & Sykes, 1977).

Given that feeling of knowing was defined with respect to memory retrieval failures, research on the topic involved creating situations where searches of memory were likely to fail. A typical paradigm was the recall–judgment–recognition procedure (Hart, 1965) in which subjects were asked general information questions and, after a failed recall and subsequent feeling of knowing rating, they were asked to choose the correct answer from a set of distractors. The calibration or validity of these feeling of knowing judgments was evaluated by comparing the ratings with the probability of selecting the correct alternative. There have been a number of extensions of this, most notably by Nelson and Narens (1980). In addition, feeling of knowing judgments have been extended to newly learned items, such as paired associates (Hart, 1967; Ryan, Petty, & Wenzlaff, 1982), and to other measures of knowing, such as relearning (Nelson, Gerler, & Narens, 1984) and perceptual identification (Nelson et al., 1984; Yaniv & Meyer, 1987).

Recently, there has been speculation that the initial evaluation of questions involves the feeling of knowing process (Nelson & Narens, 1990; Reder, 1987, 1988). Despite the research emphasis on memory retrieval failures, it seems

reasonable to conjecture that feeling of knowing is a more general process that operates whenever memory is queried. This automatic process goes unnoticed unless there is a retrieval failure after an assessment that the fact should be in memory.

There are a number of reasons to suspect that memory retrieval includes an initial evaluation of a question. First, consider the class of questions that seem unanswerable a priori, an example of which is "What was Charles Dickens's phone number?" It seems likely that people decide that they do not know the answer without first searching for the answer. Second, there is evidence that people search longer for answers to questions when they believe they know the answer (Gruneberg et al., 1977; Lachman & Lachman, 1980; Reder, 1987). Third, as mentioned earlier, there is evidence that people make strategic decisions to adopt one question-answering strategy or another without first searching memory for the answer (Reder, 1982, 1987, 1988).

The goal of the present experiments was to explore the feeling of knowing process and how it relates to strategy or procedure selection in question answering. Two important questions that have received little attention are (a) What mechanisms underlie our feeling of knowing (Nelson, Leonesio, Shimamura, Landwehr, & Narens, 1982)? and (b) What is the function of a feeling of knowing mechanism (Nelson & Narens, 1990; Reder, 1987)?

Conceivably, feelings of knowing are caused by partial retrieval of the answer to the question (e.g., letters of a trigram; Blake, 1973). Often tip-of-the-tongue states include information about the answer (e.g., the number of syllables or the first letter of the word). Some work has suggested that the feeling of knowing might be simply based on the strength of the memory trace itself (Nelson et al., 1982; Schacter, 1983), although there are some mixed results (Carroll & Simington, 1986). On the other hand, feeling of knowing might be based on something unrelated to the answer, for example the question itself. That is, it is conceivable that we typically base our feelings of knowing on familiarity judgments of the terms of the question rather than on partial retrievals of the answer. The fact that feeling of knowing in previous studies has been only weakly correlated with knowing (i.e., the ability to answer a question) suggests that the feeling of knowing that was measured might have been based on something besides knowing.

A number of researchers have suggested that feeling of knowing might be based on some form of rating of the familiarity of the question terms (Koriat & Lieblich, 1977; Metcalfe, 1986; Metcalfe & Weibe, 1987; Nelson et al., 1984; Reder, 1987). This type of mechanism predicts that feeling of knowing would correlate only moderately well with knowing, because having seen a problem or its parts would correlate only moderately well with knowing the answer. Reder (1987) showed that priming the terms in questions led subjects to have higher feelings of knowing for those questions without correspondingly raising the recall or recognition levels of the answer. In this case, judgments of question answerability may have been based on familiarity of the question. In problem solving, there is a decoupling between feeling of knowing and knowing when surface structure does not predict the deep

structure of the problem. Metcalfe (1986; Metcalfe & Weibe, 1987) has found that feeling of knowing is poorly calibrated for insight problems but is well calibrated for algebra problems.

The Game Show Paradigm

To explore the nature of feeling of knowing judgments that may occur before memory retrievals, we used a modified version of Reder's (1987) game show paradigm. In this paradigm, subjects quickly scan a question and indicate by pressing a button whether or not they think they can answer it. Their response is to be their first impression before trying to retrieve the answer. This research indicates that subjects are able to reliably estimate whether they can answer a general information question. After this quick judgment, subjects are asked to answer the question if they have estimated that they know it.[1] In the original game show experiments, subjects answered general information questions. In the present research, subjects answered arithmetic problems that were initially unfamiliar and therefore involved computation to answer (e.g., 23 * 34 = ?). Over multiple exposures to the same problem, the subject would begin to learn the answer, that is, to associate an answer with the math problem or question. This enables us to track how feeling of knowing changes with degree of learning.

We selected arithmetic problem solving as the domain because we wanted to teach facts that could be easily learned with multiple exposures to the information and facts that we could be fairly confident were not known before the experiment. In this way we could safely control the amount of exposure subjects had to the facts we wanted them to learn. We also chose arithmetic problem solving because people are distinctly aware of the two question-answering strategies, namely, retrieval of the answer and computation. This contrasts with real-world knowledge questions where people tend to be less aware of when they use retrieval and when they use plausible reasoning to answer questions (Reder, 1987).

Another reason why we selected arithmetic problem solving was that we could independently vary the frequency of exposure to (and knowledge of) the problems and the problem parts. We hypothesized that whereas the former would influence the availability of the answer, the latter would influence the feeling of knowing. Previously unstudied test problems created with parts taken from different well-studied problems may appear familiar if subjects base their decision on familiarity of the parts.

Experiment 1 was designed to address the question of whether subjects could rapidly select an appropriate strategy (retrieval or calculation) when presented with an arithmetic problem. Could subjects make such an assessment rapidly

---

[1] Interestingly, the time to decide to answer the question plus the time to come up with the answer was equal to the time to simply answer it in the condition that did not contain the two subcomponent tasks. These results suggest that some type of feeling of knowing occurs before retrieval and takes a fixed amount of time, whether it is reported or not.

with reasonable accuracy? That is, do people know quickly whether or not they know the answer to a question? Further, we hoped to learn what variables besides exposure to the answer influence this assessment. By presenting arithmetic problems multiple times and by also varying the exposure to problem parts, we hoped to be able to address these questions. Specifically, toward the end of the experiment we introduced novel problems that looked similar to studied problems, but were in fact different. Subjects were asked to judge whether they could answer arithmetic problems that they had never studied but looked like ones that they had studied. Our hypothesis was that subjects would think they could answer questions whose parts were highly familiar, even though the answer could not be known. If, on the other hand, feeling of knowing decisions are based on some implicit feedback or partial retrieval of the answer to the question, then similar-looking arithmetic problems should not result in spurious feelings of knowing.

Experiment 1

Method

Subjects. Twenty Carnegie Mellon University undergraduates participated in the experiment. Four participated for a payment of $3.50 and a monetary bonus based on their score; the remainder participated to partially fulfill a course requirement and received the same bonus.

Procedure. Subjects were told that they would be shown a large series of arithmetic problems. They sat in front of a computer screen, with a button-box and microphone on the table. After each arithmetic problem was presented on the screen, subjects rapidly indicated the strategy they would use to solve the problem; they then executed that strategy and spoke their answer into a voice-key microphone, which terminated the trial. Figure 1 shows the format for each trial.

Specifically, each trial began when the subject requested the next problem by saying "Next" to trigger the voice key. The problem was presented on a computer screen after a 500-ms delay. The subject
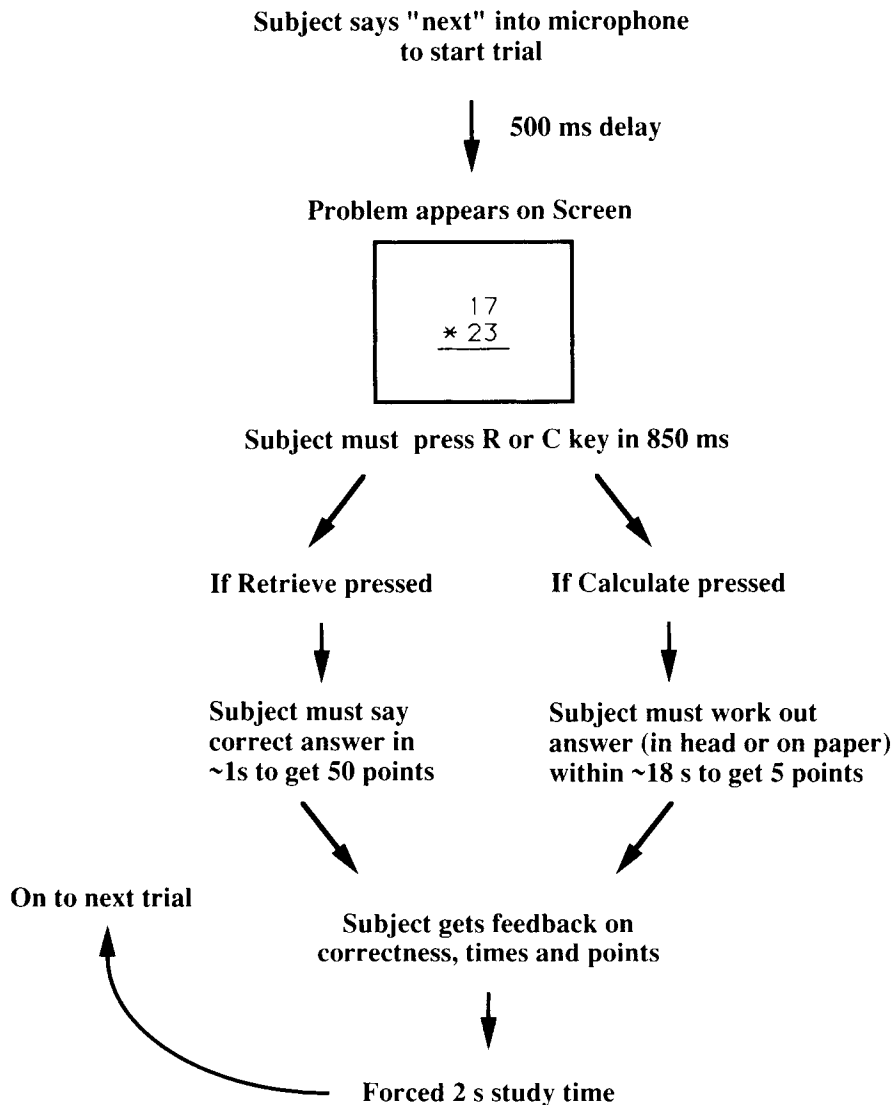

Figure 1. Procedure for a single trial. (R = retrieve, and C = calculate.)

then chose a strategy by pressing either the right button marked "R" for retrieve or the left button marked "C" for calculate. Subjects were given 850 ms to make this initial decision. This cutoff time, which was less than the minimum time required to retrieve information of comparable complexity that is not highly practiced (e.g., Millward, 1964; Staszewski, 1988), was chosen so that subjects would not base their decisions on completed retrievals instead of a feeling of knowing. The 850-ms response deadline was enforced with a large difference in points received. (Trials in which subjects took longer to choose were scored as late-to-choose responses and were analyzed separately.) The subject then either retrieved the answer from memory or calculated the answer on paper provided. The voice key recorded the onset of their answer with millisecond accuracy.

The time that subjects were allotted to answer a problem was a function of the strategy selected; the length of the correct answer; and, for calculation problems, the problem operator (addition or multiplication). When subjects selected to retrieve, they were allotted 350 ms per digit contained in the correct answer to initiate their response. When subjects chose to calculate the answer, the allotted computation time after the strategy choice was 3 s plus 3 s per digit; this calculation time limit was doubled for multiplication problems. Because all addition problems had two-digit answers, on these problems subjects were allowed 700 ms to execute the retrieve strategy and 9 s to calculate. Because multiplication problems had either three- or four-digit answers, on these problems subjects were allowed 1,050 or 1,400 ms to retrieve an answer or 24 or 30 s to calculate it. We chose these deadlines on the basis of results such as those of Staszewski (1988) with the goal of allowing time for retrieval, when that strategy was selected, but not enough time for calculation. We selected the calculation times with the view that there should be adequate time for subjects to actually perform the calculations but still be motivated to move quickly. That goal seems to have been satisfied in that far fewer than 1% of the calculation trials for either operator exceeded the deadline.

After each trial, the experimenter typed in the subject's answer or nullified the trial if the subject made a premature vocalization or failed to speak loudly enough to activate the voice key. Regardless of the subject's answer, the correct answer appeared on the screen for 2 s. Then the screen displayed the score for the current trial, the total score, the time to choose a strategy, and whether the answer was on time.

The instructions emphasized four specific payoff situations. (a) Subjects received 50 points when all of the following conditions were met: They selected the retrieval strategy; both actions, strategy choice and strategy execution, were on time; and the answer was correct. (b) Subjects received 5 points when they selected the calculation strategy and answered correctly. (c) If one of the two times (time to select the strategy or to give the answer) was late, but the answer was correct, subjects received 1 point (regardless of strategy choice). (d) When subjects met neither deadline or failed to answer correctly, they received no points. We believed that without a strong incentive to use retrieval, subjects would always choose to calculate and play it safe: At the start of the experiment all problems were new and they knew that calculation was the only viable strategy.

Subjects received 0.05 cents per point (2,000 points = $1.00). As an additional incentive to select retrieval, subjects received an extra $1 if their point total exceeded the "current highest score." The average total bonus paid was $1.79.

Thirteen practice problems were presented to familiarize subjects with the apparatus, the task, and payoff scheme. Several practice problems were duplicates of each other, and the instructions emphasized that problems would be repeated. Subjects then spent approximately 80 min completing the experiment, with a short break halfway through the problems.

Design and materials. The design varied the frequency of presentation of arithmetic problems and their parts, that is, the operands and operators. Two sets of arithmetic problems were created: training problems and posttraining problems. The training problems consisted of 216 arithmetic problems created from four different sets of 54 problems, following the design schematized in Figure 2. Each set of 54 problems contained 8 unique problems, corresponding to the eight conditions defined by a 2 × 2 × 2 factorial design. Figure 2 reflects the design of the training problems: Each level of the tree corresponds to one of the factors; the number beside a node refers to the frequency of presentation of a numeral assigned to that condition. The branches represent the levels of the three frequency factors: (a) the frequency of presentation of the top operand (high frequency in 39 problems, vs. low frequency in 15 problems), (b) the (relative) frequency of the top operand paired with a specific operator (high vs. low frequency), and (c) the frequency of presentation of the bottom operator (again, high vs. low frequency). The numbers at the terminal nodes of the tree denote how often a specific problem in that condition was presented, for example, 20 out of 54 or 3 out of 54 presentations.

The four rows of letter–operator–letter "problems" listed at the bottom of Figure 2 reflect the templates for the eight conditions. There were four repeated measures of each problem type. The assignment of numbers to these letters was random, without replacement. This random assignment was done separately for each subject, selecting from a set of 16 numbers between 14 and 36 (14, 16, 17, 18, 19, 21, 23, 24, 26, 27, 28, 29, 31, 32, 34, 36).[2]

Consider as an example the problem $a * o$ (top row, third from left). That specific problem was seen eight times during the experiment; however, the top operand a was seen a total of 39 times in this and other problems because it was a high-frequency top operand. This can be confirmed by adding up the numbers at the terminus that have an a in the first position (i.e., the first four entries: $20 + 8 + 8 + 3 = 39$). On the other hand, the $a * o$ problem was also a low-frequency first operand/operator pair. Given that a occurred 39 times in the experiment, it appeared with the multiply operator (*) only 11 times. Finally, o was a high-frequency second operand—by inspecting the four rows of problems, it can be seen that o was a second operand 39 times in the course of the training problems. To create actual problems from a template like $a * o$, we inserted randomly chosen numbers for a and o (separately for each subject). Thus, if 17 were chosen for a and 23 for o, the actual problem would be $17 * 23$.

In addition to the 216 problems just described, there were an additional 68 problems presented at the end of the experiment, each presented once. These are referred to as *posttraining* problems. Of these problems, 32 were copies of the original 32 problems presented in the training block. The remaining 36 problems were new (previously unseen during the experiment) but consisted of the same numbers and operators in new combinations. Of these new problems, 8 inverted the operands from previously presented problems (e.g., 23 * 29 would become 29 * 23), and 8 switched the operator from previously presented problems (for example, 23 * 29 would become 23 + 29). Both types of new problems (inverted problems and operator switch problems) were created from old problems from each level of original problem frequency (1, 3, 8, and 20). That is, the posttraining block included new operand reversal and operator reversal problems for problems that had been seen in their original form 1, 3, 8, or 20 times. It is important to note that operand reversals were the only new problems for which subjects were likely to know the answer.

---

[2] The excluded numbers were easier to multiply (e.g., 20) and significantly more memorable (Battig & Spera, 1962).

54
Top operand

39 High   Low   15
Operator   Operator

28 High   Low 11   11 High   Low 4
                Bottom operand

High / Low   High / Low   High / Low   High / Low

20   8   8   3   8   3   3   1

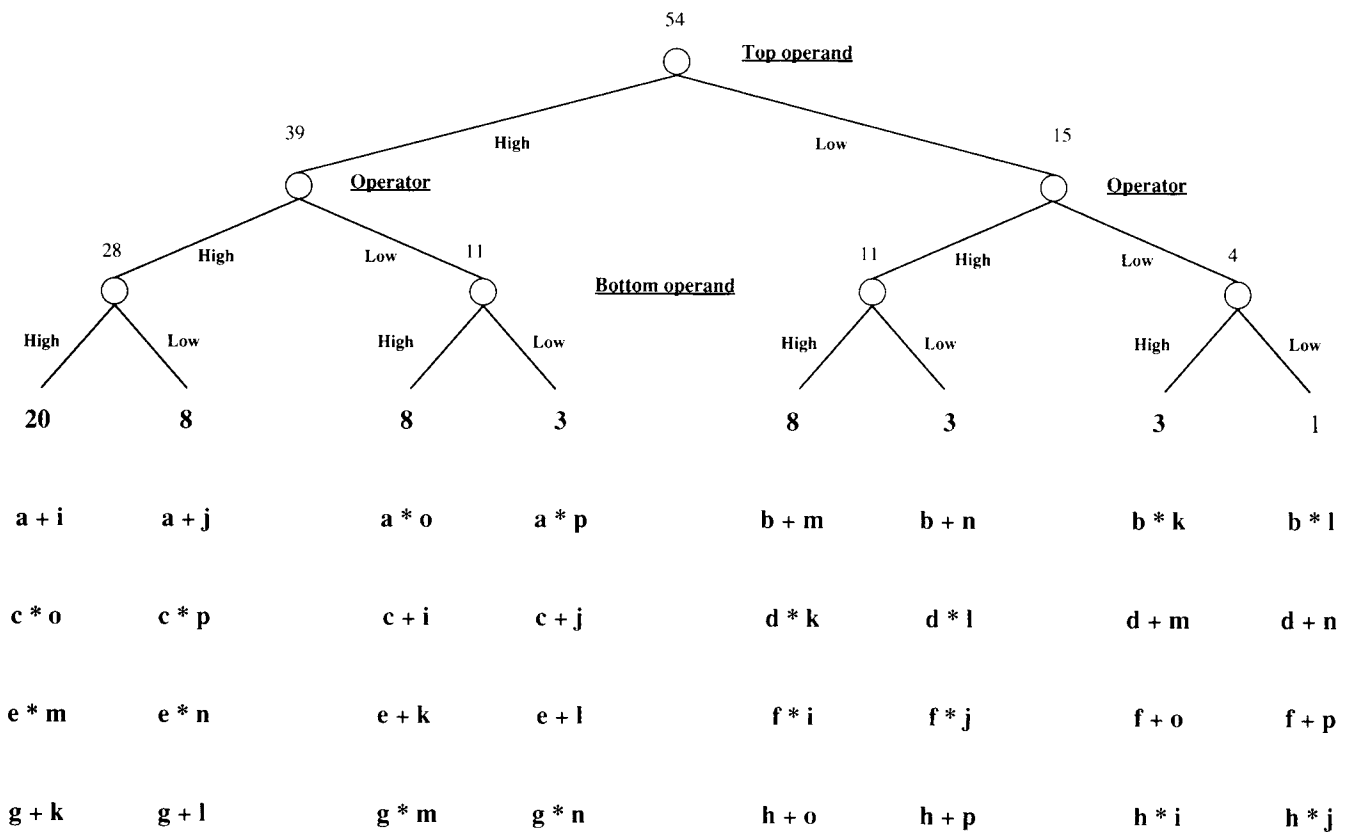| a + i | a + j | a * o | a * p | b + m | b + n | b * k | b * l |
c * o	c * p	c + i	c + j	d * k	d * l	d + m	d + n
e * m	e * n	e + k	e + l	f * i	f * j	f + o	f + p
g + k	g + l	g * m	g * n	h + o	h + p	h * i	h * j

Figure 2. Design of training problems in Experiment 1. (Left operand corresponds to top operand presented to subjects.)

The new posttraining problems also included 20 problems that repaired the operands. Thus, each of these could be classified according to the frequency of the top operand, the bottom operand, and the pairing of the operator and bottom operand. Each type of posttraining problem was presented equally often with each operator.

The 216 training problems were presented in a different random order for each subject. The 68 posttraining problems were also presented in a unique random order. Subjects were neither advised nor aware of when trials switched from the training to the posttraining set.

Results and Discussion

Less than 5% of the trials were excluded because of voice key mismeasurements of answer onset (e.g., a premature

Table 1
Subject Means for All Problems in Experiment 1

Measure	Multiplication		Addition	
	Calculation	Retrieval	Calculation	Retrieval
Strategy selected (% of trials)	81	19	31	69
Strategy selection time	685 ms	760 ms	750 ms	625 ms
Late to choose strategy (%)	19	34	33	12
Correct answer times	8,930 ms	3,660 ms	1,910 ms	780 ms
% correct	84	68	93	80
Incorrect choice of retrieval (% false alarm)	12		34	
For both strategy choices				
Gamma$_{\text{feeling of knowing and knowing}}$	.74		.76	
d'	1.96		1.29	

Note. Trials with late strategy selections were included only in the means of strategy selection time.

"umm"), and less than 1% were excluded because of very slow responding (i.e., more than 2 s to select a strategy). Analyses of variance (ANOVAs) were performed on all the dependent measures (shown by operator type in Table 1) using the factors that manipulated frequency of presentation (see Figure 2). The analyses included only data from trials in which a strategy had been chosen before the deadline. Subjects consistently chose retrieval more often for addition problems than for multiplication problems, $F(1, 19) = 38.58, p < .01, MS_e = .496$; they also were faster to select a strategy for addition problems, $F(1, 19) = 6.82, p < .02, MS_e = 11302$, and answered addition problems faster $F(1, 19) = 139.6, p < .01, MS_e = 1.934$. These differences between the operators cannot be explained simply by assuming that addition problems were learned more quickly. This can be seen from the fact that for these same addition problems, subjects also were more likely to incorrectly choose retrieval as a strategy (being either late to answer or wrong), $F(1, 19) = 35.48, p < .01, MS_e = .496$. Furthermore, on the first presentations of new addition problems, subjects attempted to retrieve the answer 58% of the time. They attempted to retrieve on only 10% of the first presentations of new multiplication problems. Indeed, retrospective protocols indicated that many subjects reported that at some point during the experiment they decided to try to earn many points and beat the system by attempting retrieval for all addition problems.

Given that many subjects treated the two operators in qualitatively different ways, we believed it was appropriate to analyze the data separately for the two operators. Our analyses focused primarily on the multiplication operator because subjects tended to treat the task in the way it was intended only for this operator; however, subjects' ability to develop the aforementioned metastrategy is interesting in its own right and is discussed further later.

The analyses of the data were organized around three questions: (a) How well could subjects perform this task? (b) How did performance change with practice? and (c) What influenced strategy choice (i.e., what variables affected subjects' beliefs that they knew the answer)?

Strategy selection time. The task required of subjects was somewhat unusual, namely, to rapidly decide which strategy—calculation or retrieval from memory—was preferable given their knowledge state. One might wonder whether subjects can actually make such a decision that is valid in such a brief period of time. There are several indications that they were able to do so. The first indication is that subjects could make a selection of a strategy quickly. Table 1 shows that the average time to select a strategy for both operators was lower than the .850-ms deadline. Overall, subjects beat the deadline in 84% of the trials. This is an underestimation of how well subjects could perform the strategy selection, because slow selection times decreased with practice: Less than 10% of the strategy choices were greater than the deadline for the last 75% of the experiment.

Accuracy of strategy choice. To successfully perform this task, subjects not only had to meet the strategy selection deadline, but also had to choose the correct strategy. By correct strategy we mean selecting retrieval if and only if the correct answer could be generated quickly. We measured accuracy both in terms of d' (Swets, 1986a, 1986b) and the

Goodman-Kruskal gamma correlation advocated by Nelson (1984, 1986).[3]

For both measures, hits were defined as trials in which subjects chose to retrieve and then correctly answered within the time allotted for retrievals; misses were defined as trials in which subjects chose to calculate and correctly answered fast enough to have met the retrieval deadline. False alarms were trials in which subjects chose to retrieve but either could not answer in time or answered incorrectly. Correct rejections were trials in which subjects chose to calculate and either did not answer within the retrieval deadline or answered incorrectly. For the d' analyses, subject conditions with no hits were assigned default Z scores of -3, and conditions with no false alarms were assigned Z scores of $+3$.

For multiplication problems, subjects had a d' of 1.97 and a gamma of .74; for addition problems, a d' of 1.29 and a gamma of .76. These measures of accuracy, especially for multiplication problems, are considered quite good. Subjects were less accurate for addition because, as mentioned, some subjects tried to calculate addition problems within the retrieval interval. That is, some subjects developed a metastrategy of hitting the retrieve button each time they saw the addition operator, regardless of whether they actually knew the answer. In subsequent analyses only multiplication problems were considered.

The effect of practice on knowing and feeling of knowing. It is interesting to compare the effects of practice on learning and on the subject's impression of knowing the information: Do both learning and feeling of knowing change at the same rate? Are they affected by the same variables in the same way? If they are, one might infer that feeling of knowing or strategy selection is based on an early read of the answer.

We used time to correctly answer a question as a measure of degree of knowing. That is, as one better learns to associate an answer with its question or problem, the response time to give that answer will decrease. In a similar fashion, we used the probability of selecting the retrieval strategy as an index of feeling of knowing. Traditionally, the term *feeling of knowing* has been used to refer to the belief that a currently irretrievable answer will be available at a later time. We are using the term in a different way. Nelson and Narens (1990) have recently used a similar paradigm, in which they asked subjects for rapid judgments of their ability to answer a question before actually answering the question. They also called these fast, first impressions "feeling of knowing."

The effect of practice on answer time. As problems were practiced, correct answer times decreased. Table 2 shows the effect of problem frequency and problem part frequency on answer time: Subjects were significantly faster when the top operand was high frequency, when the top operand/operator pair was high frequency, and when the bottom operand was high frequency: $F(1, 19) = 21.2, 39.5$, and 10.9, respectively, all $ps < .01; MS_e = 8.99, 7.93$, and 6.50, respectively. No interactions were significant (all $Fs < 3.4$). We chose to

---

[3] Gamma ranges from -1 to 1 and represents the probability that any pair of items has the same ordering in knowing (in our case, being able to correctly answer within the retrieval deadline) as it has in feeling of knowing (in our case, choosing retrieval). Both d' and gamma were computed for each subject and then averaged.

Table 2
Answer Times (in Seconds) for Correct Multiplication Problems in Experiment 1

	High-frequency top operand		Low-frequency top operand	
Bottom operand	High-frequency pairing with operator	Low-frequency pairing with operator	High-frequency pairing with operator	Low-frequency pairing with operator
High frequency	5.4 (20)	9.3 (8)	8.5 (8)	11.4 (3)
Low frequency	8.1 (8)	10.9 (3)	9.8 (3)	12.5 (1)

Note. Number of repetitions of each type of problem are in parentheses.

aggregate over trials to get more observations per condition; however, this analysis underrepresents the effect of number of presentations because the means are aggregated over all presentations, from the first to the last, in a condition. That is, the 20-presentation cell is an average of the 1st through 20th observations. If Table 2 showed only the final trials in a condition, the effects would be larger.

To get a more accurate measure of the effect of problem frequency (i.e., practice), we also regressed log answer time as a function of log practice for the most frequently presented types of multiplication problems (i.e., those that were presented 20 times). Answer time can be fit by a simple power law (Newell & Rosenbloom, 1981) based on practice with the whole problem, α (slope) $= -0.52$, $r^2 = .96$, such that the more a problem has been practiced, the faster a subject can answer it. This relationship is typical for normal learning.

The effect of practice on strategy choice. The percentage of retrievals attempted was also affected by practice. As practice with the problem increased, subjects more often chose to retrieve: A log/log regression of practice on percentage of trials selected for retrieval yielded an α of .218, $r^2 = .91$. Table 3 shows the effects of factors manipulated in the training problem set on the probability of selecting to retrieve.

Like the values in Table 2, those in Table 3 also underestimate the true effect of problem presentation frequency, for exactly the same reason. Nonetheless, there was a significant effect on tendency to select retrieval of frequency of the top operand (20.4% vs. 13.1%), an effect of the frequency of the top operand/operator pairing (22.2% vs. 11.4%), and an effect of the frequency of the bottom operand (19.9% vs. 13.7%): $F(1, 19) = 12.67, 12.49$, and 9.29, respectively, all $ps < .01$; $MS_e = 0.219, 0.462$, and 0.153, respectively. No interactions were significant (all $Fs < 1.2$).

As can be seen from Tables 2 and 3, the same factors appeared to affect knowing (as assessed by answer times) and feeling of knowing (as assessed by probability of selecting retrieval). The question remained whether knowing deter-

mined feeling of knowing or whether these two constructs were merely correlated. The answer to this question lay in an analysis of performance on the posttraining problem set. This allowed us to see whether subjects were responding to the problems as wholes or as parts.

The effect of problem subpart frequency. To test the hypothesis that strategy selection was based on the familiarity of the question parts, not the availability of the answer, we looked at performance on the novel, posttraining multiplication problem set; these problems had the property of resembling the original training problems, without having the same answers. Table 4 presents the data for both the new, posttraining problems and the copies of the original training problems that were included in the posttraining block. As mentioned earlier, there were three types of new posttraining problems: new problems made by reversing the position of the top and bottom operands, new problems made by switching the operator from addition to multiplication and vice versa, and entirely new combinations of previously presented operands. The first row of Table 4 displays the proportion of trials in which subjects selected retrieval. These values were not statistically different in an ANOVA; however, regression analyses of the data within these categories of problems (presented later) revealed evidence for effects of the frequency of the parts. (The ANOVA aggregated over subpart frequency differences.)

In Figure 3 old training problems are compared with new operator reversals on tendency to select retrieval as a function of operand pair frequency. The solid line with solid triangles represents the original, trained problems. With more exposure to the operand pairs, subjects increasingly chose to use the retrieval strategy. The slope for this function was 0.64, significantly different from 0, $t(19) = 2.29, p < .05$. For these problems, of course, exposure to the operands was confounded with practice learning the answer. The dashed line with open triangles, on the other hand, represents the operator reversal problems that had not been seen before. These prob-

Table 3
Probability of Selecting Retrieval for Multiplication Problems in Experiment 1

	High-frequency top operand		Low-frequency top operand	
Bottom operand	High-frequency pairing with operator	Low-frequency pairing with operator	High-frequency pairing with operator	Low-frequency pairing with operator
High frequency	.30 (20)	.19 (8)	.20 (8)	.11 (3)
Low frequency	.26 (8)	.06 (3)	.13 (3)	.10 (1)

Note. Number of repetitions of each type of problem are in parentheses.

Table 4

Effect of Prior Exposure on Tendency to Retrieve and Time to Retrieve or Calculate, for Multiplication Posttraining Problems in Experiment 1

Measure	Copies of training	Reversals		New combinations
		Operand	Operator	
% retrievals	13.3	17.1	11.3	9.7
% of retrievals correct and on time	21.1	0	0	0
Correct answer time (s)				
Retrievals	1.9	10.1	14.1	8.6
Calculations	8.7	10.1	10.9	10.6

lems were constructed by using operand pairs of varying frequency that had been seen on previous trials but with the other operator. Subjects could not know the answer to these problems. The slope for this line was 0.53 and does not differ statistically from the slope for the old problems ($t < 1$). Thus for both new and old problems, subjects selected retrieval in proportion to the frequency of the operand pairs, not in proportion to the frequency of exposure to the answer. We must also note, however, that the slope for the new problems was not reliably different from 0 ($t < 1$). There were very few measures per subject in these conditions, which probably accounts for this lack of power.

We also analyzed the new combination problems (operands not previously seen together) partitioned into three levels of

Figure 3. The effect of operand co-occurrence frequency on tendency to select retrieval strategy for new and old multiplication problems in Experiment 1.

frequency: those with two low-frequency operands, those with one low-frequency and one high-frequency operand, and those with two high-frequency operands. As the familiarity of the parts increased, so did the percentage of retrievals attempted: 8.3%, 8.5%, and 15%, respectively. Note that these percentages are lower than for the other new problems in which the two high-frequency operands were studied together but with a different operator (e.g., 22% for the high-frequency operator reversals vs. 15% for high-frequency new combinations).

To quantify the significance of these effects, we performed a logistic regression on strategy choice for all multiplication trials with on-time strategy selections. The predictors used were frequency of presentation of the whole problem; problem subpart frequencies, including conjoint subpart frequencies; and subject. Table 5 lists the variables used and presents their regression coefficients and corresponding Z scores. The Z scores reflect the importance of each variable for predicting strategy choice.

The negative value for the intercept constant indicates that subjects were less likely to choose retrieval overall. The range of the subject coefficients and their reliabilities reflects the substantial differences among subjects. The significant positive coefficient for the frequency with which the top and bottom operands had appeared together indicates that subjects were more likely to select retrieval when the problem parts had been seen together frequently. The coefficient for the frequency of presentation of the whole problem studied was positive but not significant. These two variables—the frequency with which the top and bottom operands had appeared together and the frequency of presentation of the whole problem—were highly correlated, because for all studied problems, these frequencies were identical. These two variables were only discriminable for the operator-switch problems. As discussed earlier, Figure 3 assures us that it was the operands appearing together that mattered, not the entire problem.[4] The other significant variable was the frequency of the bottom operand. Although this variable was significant, its regression coefficient was quite small and negative. We are somewhat puzzled as to why the value was negative, because frequency
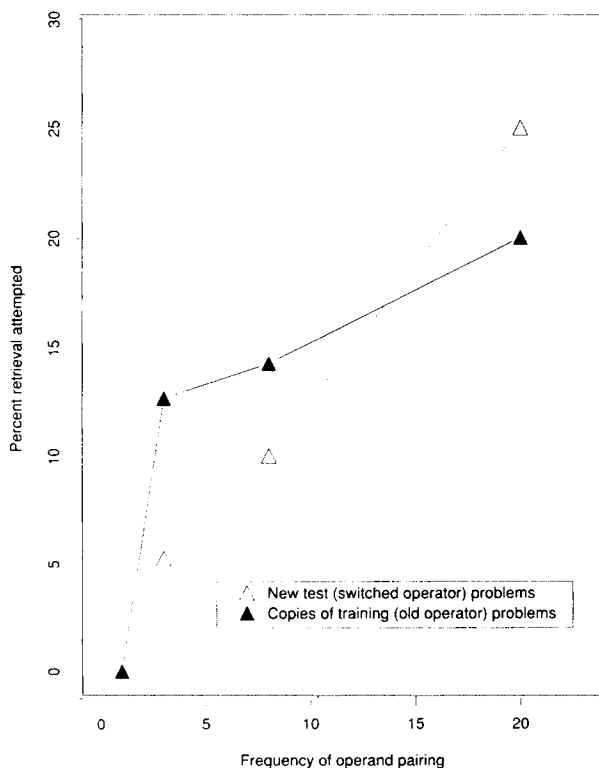
---

[4] Indeed, in a stepwise logistic regression, the frequency with which both operands appeared together had a Z score greater than 6, and the variable associated with knowing the answer, frequency of presentation of the whole problem, did not remain in the equation.

Table 5
Results of Logistic Regression on Strategy Selection for On-Time Selections of Multiplication Problems in Experiment 1

Variable	Range	Regression coefficient	$|Z|^a$
Intercept constant		−3.12	7.01**
Subject's own coded variable		−101.1 to 3.6	0–9.8
Frequency of top operand	0–48 presentations	−0.012	1.05
Frequency of bottom operand	0–48 presentations	−0.033	4.16**
Frequency with which top operand appeared with that operator	0–32 presentations	0.0007	0.03
Frequency with which both operands appeared together	0–21 presentations	0.11	2.76*
Frequency of presentation of the whole problem	0–21 presentations	0.048	1.22

[a] Computed as $(\text{coefficient} / SE_{\text{coefficient}})$ in the regression.
* $p < .05$. ** $p < .01$.

of presentation of the bottom operand and choice of retrieval correlated positively. We suspect that it compensated for a higher order nonlinear relationship between retrieval and the other variables. In any case, frequency of bottom operand had a nonsignificant, but positive, coefficient value in Experiment 2.

As an interesting aside, we should comment on the somewhat peculiar results for the operand reversal problems, for example, 34 * 23 → 23 * 34. Compared with the original problems, these showed an equivalent (or greater) tendency to be selected for retrieval, but, unlike the studied problems, were never answered correctly within the time allotted for retrieved answers. Given that multiplication is commutative, this result might seem surprising. One possible explanation is that the problem is stored in an order-specific way. Access to the learned answer requires matching this exact structure. (The feeling of knowing process that directs strategy choice does not require precise matches to this structure; rather, it can operate on partial matches to constituents.) This type of result has been seen with children learning arithmetic facts, such that the answerability of many problems is order dependent (Siegler, 1986).

Summary

In Experiment 1 we found that frequency of exposure to arithmetic problems affected both the time to answer the problems and the tendency to believe that the problems could be answered. The introduction of *posttraining* problems that were new, but contained fragments of studied problems, showed that tendency to select the retrieval strategy was influenced by familiarity with the problem parts rather than with familiarity with the answer to the question: The greater the exposure to the problem parts, the greater the tendency to select the retrieve strategy, regardless of whether the problem had been seen before.

The decision to select retrieval depended on the frequency with which the two operands appeared together, regardless of whether the problem had been seen before with that operator. It is important to note, however, that operators per se also

influenced strategy choice. Subjects tended to adopt a meta-strategy of first noting the operator of a problem. If the operator was addition, many subjects selected retrieval even if the problem seemed otherwise unfamiliar. This was because there was a strong bias in payoff to encourage retrieval. Subjects believed that they could often compute the answer to an addition problem within the time constraints for a retrieval. Although these attempts were often unsuccessful, the payoff matrix made wrong guesses not very costly, whereas risky gambles were well rewarded. In Experiment 2, this design flaw was remedied.

Experiment 2

Experiment 2 was designed to further test the hypothesis that the frequency with which the operands appeared together influenced feeling of knowing and strategy selection more than did practice with the problem or knowledge of the answer. We doubled the number of operator-switch problems in the posttraining problem set because they were the most critical for this conclusion. We also made three major changes. First, instead of using addition, we created a new operator that has a longer solution algorithm. This new function was denoted with a # (sharp). It was defined as

$$\begin{array}{ll} \text{AB} & \text{for example:} \quad 34 \\ \text{\# CD} & \qquad\qquad\quad \text{\# 56} \\ [(A + C) \ * \ (B + D) \ * \ 3] \bmod 100 & \qquad\qquad\quad 40 \end{array}$$

In order to facilitate retrieval of answers computed with this new function, the problems were designed to have a two-digit (or less) result. In order to make both operators roughly equivalent, multiplication was also redefined (and called mod-multiplication) to have a two-digit result by computing the answer modulo 100 (e.g., 34 * 29 = 86; 10 * 10 = 0).

Second, we included twin problems, problems that have the same number for both operands (sometimes referred to as tie problems). The findings from Experiment 1 suggested that subjects based strategy selection on the familiarity of the operand pair. Because there is one less number, twin problems might become especially familiar and learned particularly

quickly. Previous work has shown that this is plausible: Both adults and children are better at answering twin than nontwin problems (Ashcraft & Battaglia, 1978; Miller, Perlmutter, & Keating, 1984; Siegler, 1986). It is conceivable, however, that subjects learn to recognize twin problems faster than they learn to answer them. If so, twin problems may cause disproportionately high feelings of knowing that are erroneous.

Third, we made the problems more learnable to encourage more retrieval. In addition to using two-digit answers for multiplication and sharp, we also decreased the number of distinct problems and correspondingly increased the amount of practice with each problem.

We also changed two minor aspects of the procedure. First, in Experiment 1, our estimate of how quickly subjects could retrieve an answer was too low. So in Experiment 2 we doubled the allotted time per digit to more accurately score subjects' retrieval answer times as retrievals. Second, to encourage retrievals, we did not require a forced study time after successful retrievals but let the subjects proceed directly to the next problem. We recorded how long subjects studied a problem before proceeding.

Method

Subjects. Subjects were 14 undergraduates at Carnegie Mellon University. They received class credit or a flat fee of $5.00 for participating. As in Experiment 1, subjects also received a bonus based on performance (average bonus was $1.35).

Design and materials. The design of the training problem set was similar to the one used in Experiment 1, with several modifications. Figure 4 illustrates the three factors that were varied: the operator (* vs. #), the frequency of the top operand, and the frequency of the bottom operand. The high-frequency operands were presented twice as often as low-frequency operands, and complete problems were presented 5, 10, or 20 times. The four problem types (reflecting the four conditions of frequency) were instantiated with both operators and replicated with two sets of numbers, yielding 16 unique problems and a total of 180 problem presentations (see Figure 4). In addition, the training problem set included two twin problems that had the same number for top and bottom operand (e.g., 34 # 34). Each twin problem was presented 10 times, yielding a total of 200 training problems.

The posttraining problem set was also modified from Experiment 1 in several respects. Instead of presenting the posttraining problems after all the training problems, we introduced these special problems
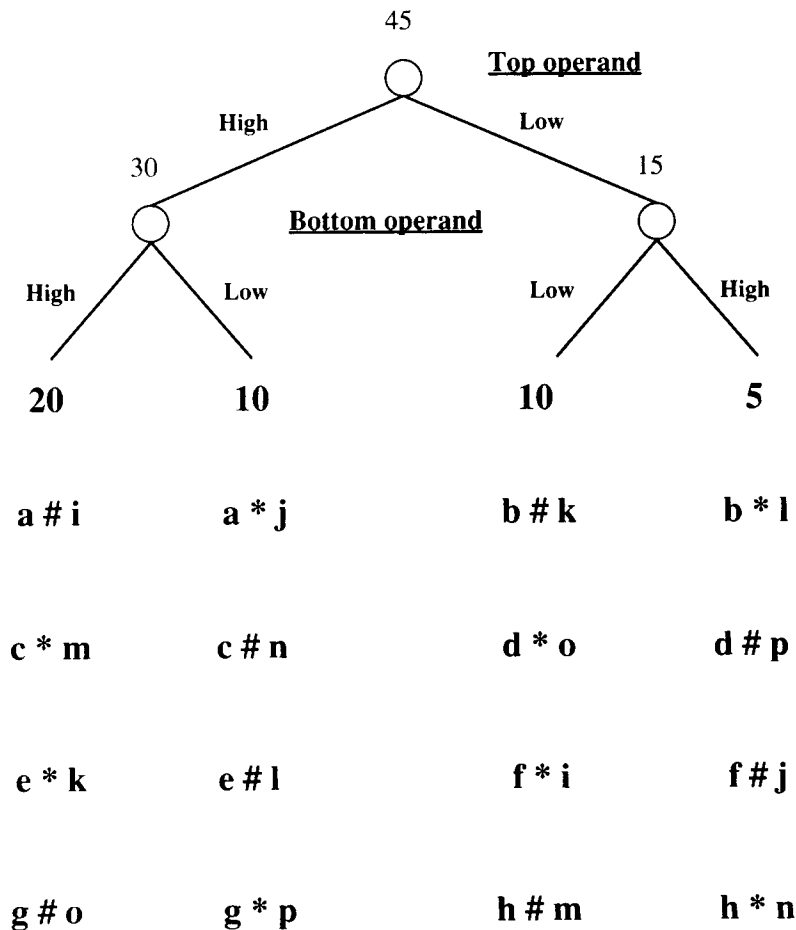

Figure 4. Design of training problems in Experiment 2.

Table 6
Subject Means for All Problems in Experiment 2

Measure	Multiplication		Sharp	
	Calculation	Retrieval	Calculation	Retrieval
Strategy selected (% of trials)	78	22	76	24
Strategy selection time	640 ms	695 ms	645 ms	580 ms
Late to choose strategy (%)	18	26	16	26
Correct answer times	9,130 ms	1,280 ms	7,850 ms	1,700 ms
% correct	85	47	80	39
Incorrect choice of retrieval (% false alarm)	13		15	
For both strategy choices				
Gamma$_{\text{feeling of knowing and knowing}}$	.76		.94	
d'	1.74		2.33	

Note. Trials with late strategy selections were included only in the means of strategy selection time.

earlier by distributing them within the second half of the training problems. Given that these problems were scattered throughout the second half of the training set, we refer to them as *nontraining* problems. There were 32 nontraining problems, of which 18 were operator-switch problems (i.e., * replaced # and vice versa), including two operator-switch twin problems. This was twice as many switch problems as in Experiment 1.

Two new types of nontraining problems were also included: First, there were eight new twin operand problems, each presented once. These were included to test whether subjects were choosing to retrieve simply for all twin problems or only for the twin problems that they had seen before. Second, there were four problems using completely new operands (i.e., operands not used in the training problem set). These were included to test the influence of extremely low frequency of presentation. These problems required adding 13, 37, 38, and 39 to the set of numbers randomly assigned to operands; otherwise, all the nontraining problems were made up of the same operands and operators used in the training problems. Each nontraining problem was presented once, with the exception of the two operator-switch twin problems, which were each presented twice.

Procedure. The procedure was the same as in Experiment 1 except that the time allowed to retrieve was doubled to 700 ms/digit (1.4 s for all answers), and subjects did not have to study the answer (for 2 s) after correct on-time retrievals. Subjects received 21 practice trials with the two operators, using different numbers to familiarize them with the procedure and new operators.

Results

Data from 1 subject were excluded from the analysis because he never chose retrieval in time. A small number of

Table 7
Answer Times in Seconds for All Correct Problems in Experiment 2

Bottom operand	High-frequency top operand	Low-frequency top operand
All trials		
High frequency	5.6 (20)	8.4 (10)
Low frequency	7.5 (10)	8.5 (5)
Last five trials		
High frequency	3.1	7.2
Low frequency	6.7	8.5

Note. Number of repetitions of each type of problem are in parentheses.

trials (2.7%) were discarded because of inaccurate voice key measurements. Also discarded were trials in which strategy selection took longer than 2 s (less than 1%).

Operator differences. Table 6 presents summary statistics for the two operators in a manner analogous to that of Table 1. Unlike addition, the sharp operator produced behavior quite similar to that produced by both multiplication and the mod-multiplication operator on all measures. It is noteworthy that operator type had no impact on tendency to select a particular strategy. In fact, the two operators in this experiment did not differ on any measure listed in Table 6 (all *F*s < 2.3), except time to correctly calculate was faster for problems with the sharp operator, $F(1, 12) = 11.31, p < .01, MS_e = .802$.[5] Therefore the data were collapsed over operator in further analyses. Although most subjects initially expressed disbelief that they could compute the sharp function, nearly all adapted to the new operator as measured by answer times, strategy choices, and accuracy.

Strategy selection time and accuracy of strategy choice. The strategy selection time results of Experiment 2 were consistent with those of Experiment 1: Subjects were able to select a strategy within the allotted time for both operators. Again, subjects were accurate in their estimation of whether they knew the answer. The average Goodman-Kruskal gamma over both operators was .92 (*n* = 12), which was comparable to the .74 in Experiment 1. The average *d'* over both operators was 2.36, and this did not differ reliably from the 1.97 in Experiment 1, $t(28) = 1.32$.

The effect of practice on answer time. As with Experiment 1, it is interesting to compare the effects of practice on learning and on the feeling of knowing, in terms of both answer time and strategy selected. As problems were practiced, correct answer times decreased. Table 7 shows the effect of problem frequency and problem part frequency on answer time averaged over all trials in the experiment and also for just the last five trials per cell. Using all trials, subjects were significantly faster when the top operand was high frequency, $F(1, 12) =$

---

[5] The faster calculation times for this novel operator can be explained by the fewer operations for sharp (two additions and three multiplications) than for multiply-mod-100 (three additions and four multiplications).

24.91, $p < .001$, $MS_e = 1.85$, and when the bottom operand was high frequency, $F(1, 12) = 6.02$, $p < .05$, $MS_e = 2.27$. The interaction term was marginally significant, $F(1, 12) = 4.40$, $p < .06$, $MS_e = 2.71$. These same effects were stronger for the last five trials: $F(1, 12) = 32.96$, $p < .001$, $MS_e = 3.24$, and $F(1, 12) = 48.35$, $p < .001$, $MS_e = 1.62$, respectively, for the top and bottom operands.[6] The interaction term was significant, $F(1, 12) = 8.47$, $p < .05$, $MS_e = 1.99$, such that subjects were not appreciably slower when both operands were low frequency.

The effect of practice on strategy choice. The percentage of retrievals attempted was also affected by practice. As practice with a problem increased, subjects more often chose to retrieve, just as in Experiment 1. Table 8 displays the probability of selecting to retrieve as a function of exposure to the problem parts. Once again, the values underestimate the differences among the conditions because they are aggregated over the first through last trial in a condition. Using all observations, an ANOVA indicated that the differences among conditions were marginally significant and in the same direction as in Experiment 1: For the top operand, $F(1, 12) = 4.67$, $p < .06$, $MS_e = .0061$; for the bottom operand, $F(1, 12) = 2.86$, $p < .15$, $MS_e = .0104$. The interaction term was not significant ($F < 2.3$).

We also conducted analyses on the last five observations in each condition. These values are displayed in the bottom half of Table 8. The effect of the frequency of exposure to the top operand and bottom operand was significant for the last five observations: $F(1, 12) = 9.3$, $p < .01$, $MS_e = .0146$, and $F(1, 12) = 5.59$, $p < .01$, $MS_e = .216$, for the top and bottom operands, respectively. The interaction term was not significant ($F < 3.2$).

Problem subpart frequency. To further test the hypothesis that strategy selection is based on the familiarity of the elements of the problem itself rather than on an early read of the answer, we once again needed only compare problems frequently seen with new problems composed of highly familiar subparts. Figure 5 shows the relevant data plotted by quartile of frequency of operand presentation. As exposure to the problem operand pairs increased with practice, subjects increasingly attempted to retrieve the answers to training problems (solid lines). Nonstudied problems (dashed lines), made by substituting the operator of a training problem, were as likely to be selected for retrieval as studied problems. The
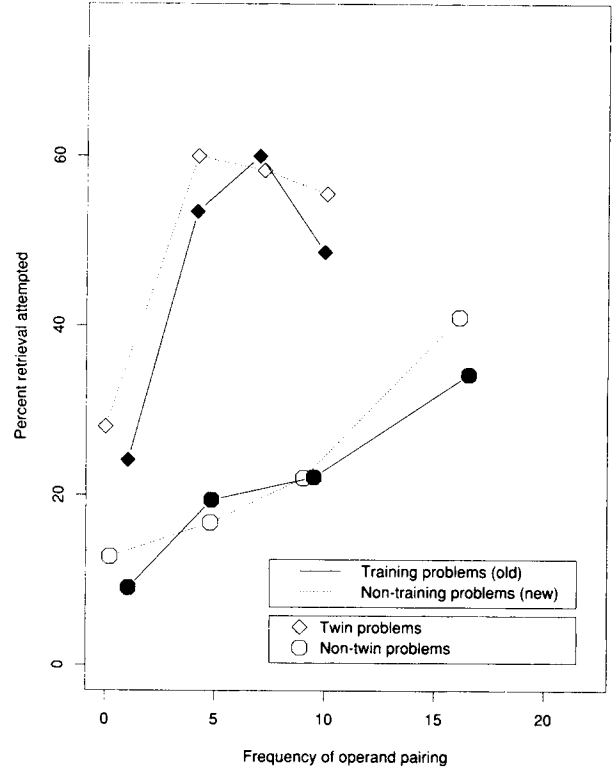

Figure 5. The effect of operand co-occurrence frequency on tendency to select retrieval strategy for twin (identical operands) and nontwin (different operands) problems in Experiment 2.

variable that seemed to matter was operand co-occurrence frequency.

The slope for new nontwin problems was 1.5, and it was 1.7 for old nontwin problems. These slopes were both reliably different from 0, $t(12) = 2.57$, $p < .05$, and $t(12) = 3.02$, $p < .05$, respectively, but were not reliably different from each other ($t < 1$). The twin problems displayed a nonlinear trend, such that tendency to select retrieve appeared to approach an asymptote for high values of operand co-occurrence frequency. We are not sure why this occurred. In any case, it might be useful to compare twin and nontwin problems on that portion of the data that appears to be linear. When we compared only the first 3 quartiles of the twin data, the slopes for new (5.6) and old (4.0) twin problems were both significantly different from 0, $t(12) = 2.63$, $p < .05$, and $t(12) = 3.01$, $p < .05$, respectively, but did not differ from each other ($t < 1$).

The other variable that influenced strategy choice was whether or not the problem was a twin (same top and bottom operand). Subjects had a stronger feeling of knowing for twins, as measured by their greater tendency to select retrieve. The

Table 8
Probability of Selecting Retrieval for All Problems in Experiment 2

Bottom operand	High-frequency top operand	Low-frequency top operand
All trials		
High frequency	.25 (20)	.18 (10)
Low frequency	.17 (10)	.15 (5)
Last five trials		
High frequency	.35	.20
Low frequency	.20	.15

Note. Number of repetitions of each type of problem are in parentheses.

___

[6] We were able to run an ANOVA on the last five observations in Experiment 2, but since some conditions in Experiment 1 had only one observation, that type of analysis was not possible for Experiment 1.

Table 9
Results of Logistic Regression on Strategy Selection for All Choice-On-Time Problems in Experiment 2

Variable	Range	Regression coefficient	$\lvert Z \rvert$[a]
Intercept constant		−4.58	10.7**
Subject's own coded variable		−0.75–6.24	0.95–12.9
Frequency of top operand	0–56 presentations	0.01	1.24
Frequency of bottom operand	0–56 presentations	0.01	0.97
Frequency with which both operands appeared together	0–20 presentations	0.11	2.96**
Frequency of presentation of whole problem	0–19 presentations	−0.01	0.44
Both operands were identical	1 = duplicated, 0 = not duplicated	1.68	5.18**
Time spent studying previous presentation of the problem	0–32.8 s	−0.02	1.44
Problem operator	1 = identical operands, 0 = nonidentical operands	−0.20	1.56

Note. [a] Computed as (coefficient / $SE_{coefficient}$) in the regression.
* $p < .05$. ** $p < .01$.

intercept for twin problems (25) significantly differed from that for nontwin problems (10), $t(12) = 2.35$, $p < .05$. The slope for twin problems (4.9) was steeper than that for nontwin problems (1.6), and this difference was marginally reliable, $t(12) = 2.08$, $p < .06$.

A logistic regression, analogous to the one performed in Experiment 1, also provided evidence for this conclusion, namely, that only familiarity with question terms mattered for the feeling of knowing. The top half of Table 9 lists the variables with their coefficients and corresponding Z scores. The significant variables were the frequency with which the two operands appeared together ($Z = 2.96$, $p < .01$) and whether or not it was a twin problem ($Z = 5.18$, $p < .01$). Variables that would be construed as influencing whether someone actually knew the answer, such as frequency of presentation of the complete problem and the amount of time spent studying the problem on its previous presentation, did not significantly influence strategy choice. In sum, just as in Experiment 1, superficial features of the problem, such as the frequency of occurrence of problem parts, affected strategy choice, whereas knowledge of the answer did not.

Twin-operand problems. We found that twin problems were more quickly learned than other problems. Figure 6 plots time of correctly answered studied problems as a function of operand co-occurrence frequency for both twins and nontwin problems. A linear regression on answer time indicated that twin problems were answered 1.5 s faster than were problems with different operands, $F(5, 2494) = 12.88$, $p < .01$. Note that this regression (and experiment) controlled for practice so that faster performance was not due to practice on the problems per se. This result indicates that twin problems are answered faster not only because they are more practiced as Siegler (1988; Siegler & Shrager, 1984) suggested, but also because they are more easily learned. Given that subjects seemed to learn twin problems faster than nontwin problems, it would be appropriate for them to choose to retrieve the answer more often for twin problems, and they did: 44% for

twin problems versus 20% for nontwin problems (see Figure 5).

It is interesting to speculate whether the greater tendency to select direct retrieval for twin problems (44% vs. 20%) was due solely to twin problems' being more easily learned (as
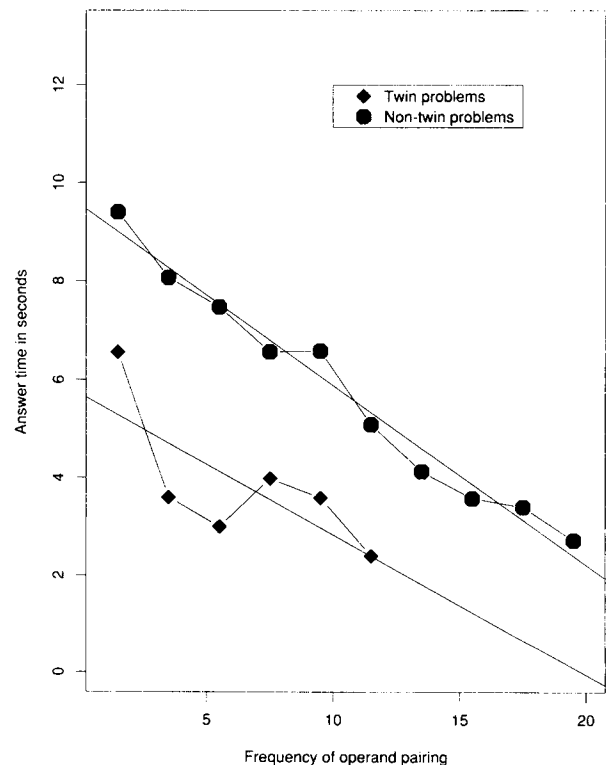

Figure 6. Correct answer times for studied twin problems and studied nontwin problems, ignoring strategy choice, as a function of operand co-occurrence frequency in Experiment 2.

shown in Figure 6) or whether there remained a bias in feeling of knowing estimates for twin problems after correcting for their faster learning rate. Figure 7 plots the propensity to select direct retrieval as a function of correct answer time for studied twin and nontwin problems. (Note that these are correct answer times regardless of which strategy was selected or whether the problem was answered on time for either strategy.) When feeling of knowing (as defined by tendency to select retrieval) was plotted against knowing (as defined by answer time), the twin and nontwin functions were on top of one another. This accuracy or calibration in feeling of knowing could also be compared using d' and gamma statistics. Faster learning for twin problems did not come at the expense of judgment accuracy: $d' = 2.53$ and gamma = .74 for twin problems; $d' = 2.17$ and gamma = .78 for nontwin problems.

General Discussion

From the present experiments, it seems clear that people are able to make an initial evaluation of a question before actually attempting to answer it. This evaluation can be used to select between alternative procedures/strategies for generating the answer. These strategy choices are made very rapidly: In both experiments, average strategy selection time was well below the deadline of 850 ms. Subjects were rarely late to choose a strategy, especially once they became accustomed to the task.
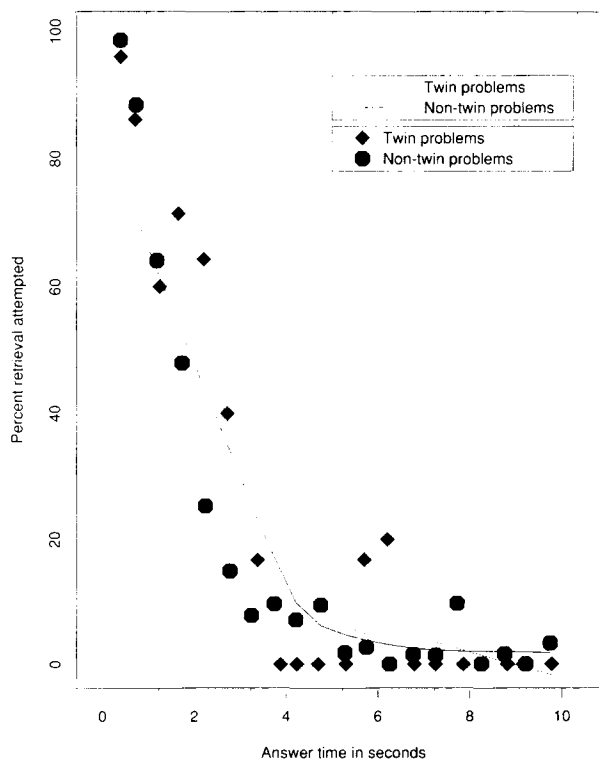

Figure 7. Tendency to select direct retrieval as a function of answer time for studied twin and nontwin problems in Experiment 2. (The points were fit using a lowess-smoothing function; Cleveland, 1979).

These rapid decisions were accurate in the sense that subjects were good at knowing which strategy was optimal; they seldom selected the calculation strategy when they knew the answer fast enough to give it within the deadline for retrieval, and vice versa. Nonetheless, this evaluation was not based on an early read of the answer: The allotted time was not enough to allow a retrieval of the answer from memory (Staszewski, 1988), and it took subjects much longer than the strategy choice deadline to retrieve the answer. Most important, subjects' strategy choice decisions were based not on how well they knew the answer, but on how familiar they were with the problem parts.

It is interesting to contrast these conclusions with findings from prior research. Others have hypothesized that feeling of knowing does come from a partial retrieval of the answer. For example, it has been found that feeling of knowing increases with the number of attributes correctly recalled (Nelson et al., 1984; Schacter & Worling, 1985). Brown and McNeill (1966) found that subjects in a tip-of-the-tongue state would often have knowledge of the first letters of the word, the number of syllables in it, and the location of its primary stress. Similarly, Blake (1973) reported that feeling of knowing increased with the number of letters correctly recalled of a paired-associate target trigram. Of course, an obvious difference between these studies and ours is that they measured feeling of knowing after a retrieval failure. In our task, subjects were required to make a very rapid decision that did not allow enough time to even attempt a retrieval. Conceivably, the very rapid judgments that we required do not share judgment processes with the more deliberate decisions in the standard feeling of knowing situation. Nonetheless, we call these quick judgments a feeling of knowing because we believe there are shared processes with the conventional task. There are only two other studies that we are aware of in which subjects displayed high accuracy for feeling of knowing judgments, one by Hart (1965) and the other by Jameson, Narens, Goldfarb, and Nelson (1990). We were able to compute gamma and d' for Hart's data (by averaging over subjects from both of his experiments) and found a gamma of .62 and a d' of 1.06. Jameson et al. obtained a gamma of .76 in one condition in which subjects were asked to predict subsequent recall. Why were subjects so much more accurate in our experiments than in many other studies that measured knowing in a variety of ways (e.g., Blake, 1973; Carroll & Simington, 1986; Gruneberg et al., 1977; Metcalfe, 1986; Metcalfe & Weibe, 1987; Nelson et al., 1984; Nelson et al., 1982)? Subjects in the game show paradigm were consistently well calibrated, with a gamma of .75 or greater.

To account for our higher accuracy in feeling of knowing, we conceive of three, nonexclusive explanations. First, the fact that our task tapped the initial, early measurement of feeling of knowing and the fact that all questions (not just failed retrievals) were rated could explain why our results are somewhat different. Because our game show paradigm required subjects to judge all questions, not just those whose answers were not recallable, we did not have a restricted range. Second, in our game show task, the terms in the question were highly correlated with the frequency of exposure to the entire question and thus should predict knowledge

of the answer. In other tasks, such as solving insight problems, this relationship explicitly does not hold. That is, the wording in a typical insight problem involves everyday objects that are not related to the structure of the problem. Third, like Jameson et al. (1990), we used recall instead of recognition as the knowledge test. Recognition may be a less sensitive measure of knowing because it may be more susceptible to plausible reasoning. Further, it may suffer from ceiling effects so that there is less room for correlations.

The present results also contrast with those on comprehension monitoring (e.g., Glenberg & Epstein, 1985; Glenberg, Sanocki, Epstein, & Morris, 1987). In comprehension-monitoring tasks, subjects are asked to estimate how well they understand the presented material and to predict how well they will do on an exam. Subjects' calibration measures are extremely poor. Glenberg (1991) explained this surprisingly low performance by noting that subjects would have to be able to predict the questions to be asked. In our task, subjects were estimating their knowledge of a question they had already seen, if only briefly.

Findings of several previous studies are consistent with the view that the feeling of knowing is caused by a feeling of familiarity with the question itself, rather than knowledge of the answer. Reder (1987) found that subjects erroneously believed they could answer general information questions when the questions contained terms that had been primed earlier in another part of the experiment. Koriat and Lieblich (1977) found that repeating the question (without the answer) increased people's feeling of knowing. In contrast, Jameson et al. (1990) primed the answers to the questions rather than priming terms of the question. They found that the priming increased the availability of the answers but did not influence feeling of knowing!

There is other independent evidence that feeling of knowing judgments are related to question-answering behavior; for example, high feelings of knowing correlate with longer search times (Gruneberg et al., 1977; Lachman & Lachman, 1980; Nelson et al., 1984; Reder, 1987, 1988; Ryan et al., 1982). Similarly, Costermans, Lories, and Ansay (1992) found that ratings of the feeling of knowing for questions whose answers could not be recalled were not good predictors of subsequent recognition of the correct answer. On the other hand, these ratings were correlated with search times for trying to find the answer, again suggesting that subjects use these judgments to guide their question-answering strategies even though they do not have access to the answer.

How might a feeling of knowing mechanism work? Reder (1987) suggested that initial evaluation of a question might involve determining (a) how familiar terms in the question seem and (b) how much knowledge seems related to the question. Familiarity was operationalized as the activation level of a term "relative to its base activation level" (Reder, p. 121). When a question is asked, all terms in the question are raised to the same high level of activation and placed into working memory. These terms have immediate access to their corresponding nodes in long-term memory.

The critical comparison is the current activation level of the term's long-term memory counterpart. Base or generic activation levels of low-frequency words tend to be lower than those of high-frequency words. Therefore, if a low-frequency word has been seen recently, the change in activation is more salient than it is for a high-frequency word. As activation decays, the change will be noticeable longer for low-frequency words.

In the present experiments, all of the numbers in the arithmetic problems had been seen by subjects a lot recently. Therefore, mere familiarity (change in base activation level) would not be enough to give a feeling of knowing. We speculate that in a situation such as this, intersection of activation is critical. This intersection occurs when the terms in the question contact their corresponding nodes in long-term memory and these nodes send out activation along relatively strong association paths (i.e., activation in proportion to interassociation strength). If the terms have appeared together often in recent history, the activation will intersect (see Anderson, 1983). The compound-cue theory of Ratcliff and McKoon (1988) could probably be used to make a similar model.

This view is supported by our finding that co-occurrence of the two operands was a much better predictor of the tendency to select retrieval than was frequency of the individual components. Other research is consistent with this view, for example, studies by Dosher and Rosedale (1989, 1990). They found that when a word from a studied triple was presented as a prime for recognition of the remaining two words, recognition was improved only when all three were studied together. Furthermore, Dosher and Rosedale (1990) showed that when a test word was independently cued with two prime words, each of which was directly studied with the test item, priming was significant only if all three went together in a studied triple. This is consistent with our view that co-occurrence information, not individual associations, are important for giving a feeling of knowing. Dosher and Rosedale were not using a feeling of knowing paradigm; however, they were examining the time course of priming (i.e., very early effects of priming). Their results support the view that there is rapid accessibility to the representation of an ensemble, that is, items that co-occur. In contrast, they found that other partial associations primed only weakly if at all.

It is important to distinguish between retrieval processes and feeling of knowing processes. The former are not automatic, consume mental resources, and vary in response time to complete. The initial feeling of knowing explored in the present experiments is thought to proceed rapidly with minimal effort. Unlike retrieval, it does not require careful inspection of the memory traces. In fact, our results show that the time to select a strategy was affected by practice with the task, but not by practice with a given problem. That is, as subjects became familiar with the somewhat unusual task, they got faster at pushing a button to reflect their choice; yet this judgment time was unaffected by (a) whether they chose to retrieve or to calculate, (b) how quickly they could answer the question, or (c) how many times they had seen the answer. This result is consistent with the view that this judgment comes from a processing stage that is independent of the retrieval process itself.

In conclusion, the present results suggest that in everyday situations, an initial feeling of knowing can be used to select

appropriate question-answering strategies. Although in every-day situations one is not consciously aware of making fast decisions about knowing, we believe that such judgments occur implicitly whenever a question is asked. An exaggerated form of this feeling of knowing occurs in nonlaboratory settings when people are contestants on television game shows, trying to estimate whether they can answer a question (posed also to an opponent) before it is heard in its entirety. In sum, we believe that this fast, initial judgment is based on surface features of the question and that, in most instances, this is a useful *heuristic*. As we have shown, fast judgments based on features of the question are highly calibrated except when experimenters intentionally mislead the subject.

References

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Ashcraft, M. H., & Battaglia, J. (1978). Cognitive arithmetic: Evidence for retrieval and decision processes in mental addition. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 527–538.

Battig, W. F., & Spera, A. J. (1962). Rated association values of numbers from 0–100. *Journal of Verbal Learning and Verbal Behavior, 1,* 200–202.

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

Carroll, M., & Simington, A. (1986). The effects of degree of learning, meaning, and individual differences on feeling-of-knowing a word's meaning. *Acta Psychologica, 61,* 3–16.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of American Statistical Association, 74,* 829–836.

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.

Dosher, B. A., & Rosedale, G. (1989). Integrated retrieval cues as a mechanism for priming in retrieval from memory. *Journal of Experimental Psychology: General, 118,* 191–211.

Dosher, B. A., & Rosedale, G. (1990, November). *Dual cues in item recognition: An ensemble model of priming.* Paper presented at the Thirty-First Annual Meeting of the Psychonomics Society, New Orleans, LA.

Glenberg, A. (1991, July). *The knowledge used in calibrating comprehension.* Paper presented at the International Conference on Memory, Lancaster, England.

Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 702–718.

Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General, 116,* 119–136.

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

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,* 1–14.

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

Lachman, J. L., & Lachman, R. (1980). Age and the actualization of world knowledge. In L. W. Poon, J. L. Fozard, L. S. Cermak, D. Arenberg, & L. W. Thompson (Eds.), *New directions in memory and aging* (pp. 313–343). Hillsdale, NJ: Erlbaum.

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

Metcalfe, J., & Weibe, D. (1987). Intuition in insight and noninsight problem solving. *Memory & Cognition, 15,* 238–246.

Miller, K., Perlmutter, M., & Keating, D. (1984). Cognitive arithmetic: Comparison of operations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 46–60.

Millward, R. (1964). Latency in a modified paired-associate learning experiment. *Journal of Verbal Learning and Verbal Behavior, 3,* 309–316.

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. (1986). ROC curves and measures of discrimination accuracy: A reply to Swets. *Psychological Bulletin, 100,* 128–132.

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

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 1–45). New York: Academic Press.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–51) Hillsdale, NJ: Erlbaum.

Norman, D. A. (1973). Memory, knowledge, and the answering of questions. In R. H. Solso (Ed.), *Contemporary issues in cognitive psychology: The Loyola Symposium* (pp. 135–165). Washington, DC: Winston.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review, 95,* 385–408.

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

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology, 19,* 90–137.

Reder, L. M. (1988). Strategic control of retrieval strategies. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 227–259). San Diego, CA: Academic Press.

Ryan, M. P., Petty, C. R., & Wenzlaff, R. M. (1982). Motivated remembering efforts during tip-of-the-tongue states. *Acta Psychologica, 51,* 137–147.

Schacter, D. L. (1983). Feeling of knowing in episodic memory. *Canadian Journal of Psychology, 9,* 39–54.

Schacter, D. L., & Worling, J. R. (1985). Attribute information and the feeling-of-knowing. *Canadian Journal of Psychology, 39,* 467–475.

Siegler, R. S. (1986). *Children's thinking*. Englewood Cliffs, NJ: Prentice-Hall.

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General, 117*, 258–275.

Siegler, R. S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Eds.), *Origins of cognitive skills* (pp. 229–293). Hillsdale, NJ: Erlbaum.

Staszewski, J. J. (1988). Skilled memory and expert mental calculation. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 71–128). Hillsdale, NJ: Erlbaum.

Swets, J. A. (1986a). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin, 99*, 181–198.

Swets, J. A. (1986b). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin, 99*, 100–117.

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.