

To Calculate or Not to Calculate: A Source Activation Confusion Model of Problem Familiarity's Role in Strategy Selection

Christian D. Schunn, Lynne M. Reder, Adisack Nhouyvanisvong,
Daniel R. Richards, and Philip J. Stroffolino
Carnegie Mellon University

How do people decide whether to try to retrieve an answer to a problem or to compute the answer by some other means? The authors report 2 experiments showing that this decision is based on problem familiarity rather than on retrievability of some answer (correct or incorrect), even when problem familiarization occurred 24 hr earlier. These effects at the level of the individual problem solver and the results reported by L. M. Reder and F. E. Ritter (1992) are well fit with the same parameter values in a spreading-activation computational model of feeling of knowing in which decisions to retrieve or compute an answer are based on the familiarity or activation levels of the problem representation. The authors therefore argue that strategy selection is governed by a familiarity-based feeling-of-knowing process rather than by a process that uses the availability of the answer or some form of race between retrieving and computing the answer.

When given any problem to solve, the problem solver may choose to retrieve a previously computed solution from memory or choose to compute the answer by using some reasoning strategy (e.g., using an algorithm, inferencing, or making plausibility judgments). This decision between retrieving and computing is used in a wide range of problem domains. In academic domains, the decision is important in tasks varying in complexity from simple arithmetic (e.g., $9 + 6$) to fact verification in story comprehension (e.g., Did the heir to the hamburger chain love his wife?) to economics (e.g., What is the effect of a value added tax on supply and demand?). In everyday problem domains, the decision is also important in a wide range of tasks such as navigating a path to the grocery store or calling someone on the phone.

How is the decision between retrieval and reasoning

Christian D. Schunn, Lynne M. Reder, and Adisack Nhouyvanisvong, Department of Psychology, Carnegie Mellon University; Daniel R. Richards and Philip J. Stroffolino, Department of Computer Science, Carnegie Mellon University. Daniel R. Richards is now at the Department of Behavior Genetics, Stanford University; Philip J. Stroffolino is now at Maya, Pittsburgh, PA.

Portions of the experiments and simulations were also described in Miner and Reder (1994) and in Reder and Schunn (1996). The original simulations were developed by Philip J. Stroffolino. Experiment 1 was conducted by and the original model fits were done by Daniel R. Richards.

The work reported in this article was supported by graduate fellowships from the Natural Sciences and Engineering Research Council and la Formation des Chercheurs et l'Aide à la Recherche and by Grant BNS-8908030 from the National Science Foundation and Grant N00014-95-1-0223 from the Office of Naval Research. We thank John Anderson and Marsha Lovett for their helpful comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Christian D. Schunn or Lynne M. Reder, Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213. Electronic mail may be sent via Internet to Christian D. Schunn at schunn@cmu.edu or to Lynne M. Reder at reder@cmu.edu.

made? A simple model of cognition is one in which retrieval is used when the answer is known, and other strategies are used when the answer is not known. There are many variants of this simple model. The chunking algorithm used in the Soar cognitive architecture (Laird, Newell, & Rosenbloom, 1987; Newell, 1990) provides a simple example. When a new problem is solved, the chunking algorithm creates a rule that will retrieve the solution should that same problem reoccur. Another variant of this type of model asserts that retrieval is always attempted first, and a calculation strategy is attempted only when the retrieval process either fails to produce any answer within a given time limit or fails to produce an answer of sufficient strength (e.g., Siegler, 1987, 1988; Siegler & Shrager, 1984). A third variant assumes that there is a race between the process of retrieving an answer and computing the answer (e.g., Logan, 1988). If the answer can be retrieved before an answer can be computed, then the retrieved answer is used; otherwise the answer is computed.

However, there are several empirical phenomena that are problematic for these simple models. Reder (1982, 1987, 1988) presented a strategy selection model and empirical data that strongly rejected the notion that people will always retrieve the answer if it is known. These experiments showed that question-answering strategy choice can be influenced by factors of the questioning situation and the question itself. For example, Reder (1987) found that participants' tendency to search for an answer rather than compute an answer was affected by variables such as prior history of success with retrieval on other problems, specific advice as to which strategy was more likely to work, and whether components of the question had been primed earlier.

How might these factors influence the decision to retrieve or compute? A general process that has been argued to underlie this decision is *feeling of knowing* (e.g., Miner & Reder, 1994; Nelson & Narens, 1990; Reder, 1987, 1988).

Feeling of knowing is the degree of belief that a piece of information can be retrieved from memory. Although some researchers have focused on feeling-of-knowing judgments made after memory retrieval failures (e.g., Connor, Balota, & Neely, 1992; Gruneberg & Monks, 1974; Hart, 1965; Nelson & Narens, 1980; Schwartz & Metcalfe, 1992; Yaniv & Meyer, 1987), others have conceptualized feeling of knowing as a rapid, preretrieval process (e.g., Miner & Reder, 1994; Reder, 1987, 1988; Reder & Ritter, 1992; Schreiber & Nelson, 1995). It is this rapid feeling-of-knowing process that we argue is being used in the decision to retrieve or compute.

For example, in an arithmetic task, Reder and Ritter (1992) studied the process by which individuals decide whether to retrieve or calculate. The paradigm they used had participants deciding whether to retrieve or compute within 850 ms after being presented 2 digit \times 2 digit multiplication and addition problems. This short deadline was used to ensure that participants could not retrieve the answer before making the decision. After the decision, participants gave their answer, either by retrieving the answer from memory or by computing the answer. The amount of time given to participants to respond depended on their initial decision. If they chose to retrieve, then they were given 1,400 ms to initiate their response. If they chose to calculate, then they were given 20 s. A payoff scheme was used that heavily rewarded correct, on-time retrievals.

Using this paradigm, Reder and Ritter (1992) found that participants' retrieve-compute decisions were quite accurate. That is, participants were usually able to retrieve the answer when they chose to retrieve, and they were usually not able to retrieve when they chose to compute. Thus, participants were quite able to make these decisions while selecting quickly. Moreover, Reder (1987) found that the time to simply answer the question without first explicitly reporting the strategy choice was equal to the time to report the strategy choice plus the time to then give an answer. This finding is consistent with the claim that the strategy choice process is a natural part of the problem-answering process.

What mechanism subserves this rapid feeling-of-knowing process? Reder and Ritter (1992) suggested that these decisions were not based on an early read of the answer. Rather, they argued that the decisions were based on familiarity with the problem. In support of this argument, they found that familiarity with components of the problem strongly predicted participants' feeling-of-knowing judgments, most notably in the case of novel problems that were similar to previously seen problems. For example, some of the previously seen problems were presented with the operators switched (i.e., multiply instead of add or vice versa). Despite not knowing the answers to these operator-switch problems, participants were just as likely to think they knew the answer to these problems as they were to the original problems. Similarly, Reder (1987) found that surreptitious familiarization with words in a question also led to spurious feelings of knowing. Recently, other researchers have also found evidence suggesting that feeling-of-knowing judgments are based on features of the problem rather than on features of the solution (e.g., Connor et al.,

1992; Schreiber & Nelson, 1995; Schwartz & Metcalfe, 1992).

Thus, existing research suggests that individuals decide whether to retrieve before attempting a memory retrieval. However, there are several potential problems with this account. First, there is an alternative account of the previous findings that may be consistent with the theory that participants always try to retrieve first. Under this account, participants may be basing their feeling-of-knowing judgments on an early read of some answer, not necessarily the correct answer. That is, in previous experiments, when participants have been fooled into believing that they know the answer when in fact they have only been exposed to similar items (e.g., Reder & Ritter, 1992), they may have based their responses on partial retrievals of the answer to those similar problems (i.e., the wrong answer). Thus, while Reder and Ritter argued that their results provided evidence for a preretrieval strategy selection process based purely on the familiarity of the problem statement (and not on any aspect of the answer), it may be that participants actually always do attempt to retrieve first and that the feeling-of-knowing judgments simply reflect the initial, potentially incorrect, outcomes of that retrieval process. In support of this interpretation, Schwartz and Metcalfe (1992) found in some of their experiments that priming the answer occasionally increased feeling-of-knowing judgments and that priming the cue occasionally decreased accuracy of recall. The equivocal nature of the Schwartz and Metcalfe results does not bear on the issue of strategy selection because they only studied feeling of knowing after recall failures (i.e., well after strategy decisions had been made); however, the alternative interpretation of the Reder and Ritter results is potentially damaging to an account of strategy selection that is based on familiarity.

The second potential problem with the role of feeling of knowing in strategy choice relates to potential external validity problems with the paradigm used to establish this connection. In previous experiments (e.g., Reder & Ritter, 1992; Schwartz & Metcalfe, 1992), participants gave feeling-of-knowing judgments only for items that they had recently encountered (i.e., during the experimental session). In other words, feeling-of-knowing judgments were made after relatively short delays from the last exposure. It may be that under these artificially short delays, participants made feeling-of-knowing judgments on the basis of still active problem representations. Further, under longer delays, it may be that people are not able to use the relatively lower, long-term activation levels of problem representations to guide strategy choice.¹ Thus, the previous findings might not generalize to the more typical, long-term case.

Third, the use of feeling of knowing to decide whether to retrieve or compute presents a computational conundrum: What is the advantage of first computing feeling of knowing over simply attempting to retrieve immediately? That is, why is computing feeling of knowing easier than retrieving the answer? If they were equally difficult, there would be no

¹ We thank Robert Siegler for pointing out this possibility.

advantage of computing feeling of knowing before retrieving over always trying to retrieve first. This computational conundrum may be why most models of strategy choice have assumed that participants always try to retrieve first.

In summary, there are several potentially valid alternative explanations questioning the conclusion that individuals decide whether to retrieve or compute, and there are computational issues suggesting that individuals always retrieve first. To address these issues, we present a new computational model of feeling of knowing and two new experiments. We begin with an experiment that establishes the phenomenon that is to be modeled. We then turn to a description of the model and provide detailed fits of the model to the data at an individual participant level. Experiment 2 is then presented as further test of the model, as well as an extension of the generality of the basic empirical findings.

Experiment 1

In Experiment 1 we sought to test one of the alternative explanations for the findings of previous studies that argued for the problem-familiarity hypothesis. Under this explanation, participants may not have been making retrieve-compute decisions on the basis of familiarity with the problem. Instead, participants might actually have made these decisions on the basis of an early read of an answer. Although previous work using similar-looking trick problems has established that participants' decisions were not based on an early read of the correct answer to a particular problem, participants might have misrecognized the trick problems as the original problems and based their decisions on an early read of the (now incorrect) answer to the original problem. In Experiment 1 we attempted to eliminate this alternative explanation, and we used a variant of the rapid feeling-of-knowing decision paradigm used by Reder and Ritter (1992).

As in Reder and Ritter (1992) participants in Experiment 1 were repeatedly presented with arithmetic problems and were asked to make a rapid decision as to whether they would retrieve or compute the answer. However, to rule out the explanation that participants were making retrieve-compute decisions on the basis of early read of some answer, special problems were created for which participants made the initial retrieve-compute decisions but, on most of the trials for these problems, they were not allowed to actually retrieve or compute an answer. On these special trials, the screen was cleared immediately after the retrieve-compute choice, and the participant was instructed to continue onto the next question. Occasionally, participants were required to give an answer to the problems assigned special status. This was done to allow problem strength and answer strength to vary somewhat independently and to ensure that participants could not learn that these problems were never answered. Only a subset of the problems was assigned to this special status—other problems were answered whenever they were seen. Thus, for these special problems, called *infrequently answered problems*, problem familiarity was increased, but the associated answers were typically not

strengthened. If the early-read hypothesis is correct, then participants should not select retrieve for these infrequently answered problems because the answer is not being associated with them on most of the trials. However, if the problem-familiarity hypothesis is correct, then participants should select retrieve for these infrequently answered problems, and the probability of selecting retrieve should be a simple function of the amount of exposure to the problems because the familiarity of the problem is being increased with each exposure.

Method

Participants. Twenty-five Carnegie Mellon University and University of Pittsburgh graduates and undergraduates participated in the experiment. They were paid on the basis of their performance. However, the minimum payment was \$5/hr.

Procedure. Participants were told that they would answer a large series of arithmetic problems. Half of the problems involved multiplication, and half involved an invented operator (described below). Different problems were presented at different frequencies in a random order. Thus, as the session progressed, answers to frequently presented problems would be learned.

The participants sat in front of a computer monitor with a button box and microphone on the table. After each arithmetic problem was presented on the screen, participants rapidly chose to either calculate or retrieve the answer with the button box; they then executed their strategy and spoke their answer into the voice key microphone, which ended the trial.

Specifically, each trial began when the participant said "next" and triggered the voice key. The problem was displayed 0.5 s after the trigger, in large font on the screen. The operands were presented vertically, one on top of the other, and the operator was presented to the left of the bottom operand. The participant then chose a strategy by pressing either the right button marked *R* for retrieve or the left button marked *C* for calculate. Participants were given 850 ms to make this initial decision. The 850-ms response deadline was enforced with a large difference in points received. A letter prompt (*R* or *C*) indicating the participant's decision was displayed on the screen. The participant then either retrieved the answer from memory or calculated the answer on paper. The voice key recorded the onset of the answer with millisecond accuracy.

The time that participants were given to answer a problem was a function of the strategy selected. When participants chose to retrieve, they were then allotted 1.4 s to initiate their response. By contrast, when participants selected calculate, they were given 20 s to compute, before they had to initiate their response. Both operations (multiplication, $*$, and sharp, $\#$) were modulo 100, so only the right-most two digits were given. For a problem $ab \# cd$, sharp was defined as $[(a + c) * (b + d) * 3]$ modulo 100. For example, $52 \# 34 = [8 * 6 * 3]$ modulo 100 = 144 modulo 100 = 44. In addition to being given the definition of the sharp operator before starting the task, participants were also given several practice problems.

The retrieve cutoff was chosen to allow sufficient time to retrieve but not enough time to calculate the answer. The calculation cutoff was selected to provide sufficient time for participants to perform the calculations but provide motivation to work quickly. That goal seems to have been satisfied in that fewer than 1% of the calculation trials for either operator exceeded the deadline.

After each trial, the experimenter typed in the participant's answer or nullified the trial if the participant made a premature vocalization or failed to speak loudly enough to activate the voice key. Then the screen displayed the score for the current trial, the

total score, the time to choose a strategy, whether the strategy choice or answer was late, and the correct answer, which always remained on the screen until the participant requested the next question by speaking into the microphone.

One fourth of the trials involved infrequently answered problems. Each infrequently answered problem was answered only two out of every seven times it was presented. On each of the other five of seven trials, when the participant was not asked to answer the problem, the screen was cleared immediately after the retrieve-compute choice, and the participant was instructed to continue onto the next question.

The instructions emphasized five specific payoff situations: (a) Participants received 50 points when they selected the retrieval strategy, and both strategy choice and strategy execution were on time, and the answer was correct (i.e., all three conditions were satisfied); (b) participants received 5 points when they selected the calculation strategy and met the comparable constraints; (c) if one of the two times (time to select the strategy or to give the answer) was late, but the answer was correct, participants received 1 point (regardless of strategy choice); (d) for infrequently answered problem trials that were not answered, participants received 2 points, regardless of their strategy choice; and (e) when participants met neither deadline or failed to answer correctly, they received no points. At the start of the experiment, all problems were new and the participants understood that calculation was the only viable strategy. Without a strong incentive to use retrieval, participants would play it safe and always select to calculate.

Participants received 0.1 cents per point (1,000 points = \$1.00). As an additional incentive to select retrieval, participants received an extra dollar if their point total exceeded the current highest score. The average total paid was \$7.71.

Twenty-one practice problems were presented to familiarize participants with the apparatus, the task, and the payoff scheme. Several practice problems were duplicates of each other, and the instructions emphasized that problems would be repeated and that some would not be answered. Participants then spent approximately 75 min completing the experiment, with a short break halfway through the problems.

Design and materials. Presentation frequency of problems varied two factors: the frequency of the top operand, and the frequency of the bottom operand. This design was replicated for both operators (* vs. #). Two sets of arithmetic problems were created consisting of 12 normal problems and 4 infrequently answered problems. There were also several frequency levels for problems of each type. The design is illustrated in Figure 1. Each level of the tree corresponds to one of the two factors; the number inside the node refers to the frequency of presentation of a numeral assigned to that condition. The branches represent the levels of the two frequency factors: (a) the frequency of presentation of the top operand (a high-frequency operand occurred in 42 problems, whereas a low-frequency operand occurred in 21 problems) and (b) the frequency of presentation of the bottom operand (again, high vs. low frequency). The numbers at the terminal nodes of the tree indicate how often a specific problem in that condition was presented (e.g., 14 presentations for the high-low frequency pairs). Complete problems were presented 7, 14, or 28 times, producing a total of 189 normal problem trials and 63 infrequently answered problem trials.

The four rows of letter-operator-letter problems listed at the bottom of Figure 1 represent the complete set of problems presented to a given participant. Each letter corresponds to a number. Note that any given number appears in only two problems: one multiply (*) problem and one sharp (#) problem. The assignment of numbers to these letters was random, without replacement. This random assignment was done separately for each participant,

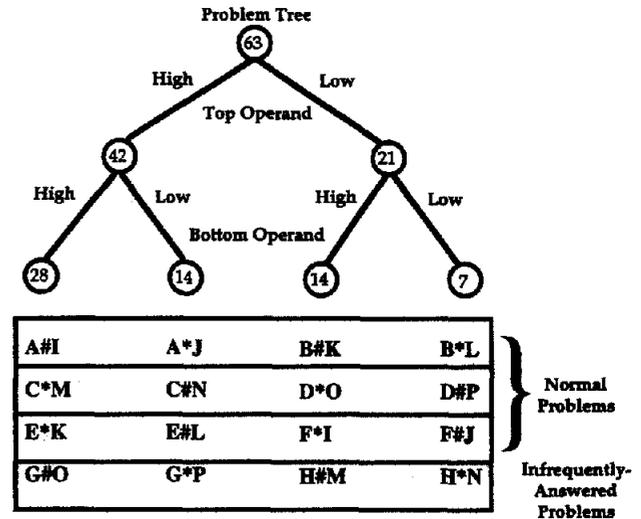


Figure 1. The problem frequency template for Experiment 1. The four rows of letter-operator-letter problems represent the complete set of problems presented to a given participant. Each letter corresponds to a number. Any given number appears in only 2 problems: one multiply (*) problem and one sharp (#) problem.

selecting from a set of 16 numbers between 14 and 38 (14, 16, 17, 18, 19, 23, 24, 26, 27, 28, 29, 32, 34, 36, 37, 38). The excluded numbers were easier to multiply and more memorable (Battig & Spera, 1962). The particular row in the template that was assigned to infrequently answered problems was varied across participants (i.e., it was not always the fourth row as is suggested by the figure). Thus, across all participants, infrequently answered problems appeared with both operators equally often.

The presentation order of the normal problems was randomly determined for each participant. The 63 infrequently answered problem trials were also distributed randomly within the normal problem trials. Again, the infrequently answered problems differed from the normal problems in that, on average, 5 of every 7 infrequently answered problem presentations were not to be answered. However, because the ordering of infrequently answered problems was random for each participant, the number of infrequently answered that was actually answered in a given string of 7 such problems could have been more or less than 5. Because the presentation order of the normal and infrequently answered problems was completely random, and because the problems varied in their overall presentation frequency, the number of times that a given problem had been seen previously was not strongly correlated with trial number. This complex design prevented participants from being able to use a simple strategy of simply selecting retrieve gradually more often over the course of the task without paying attention to particular problem characteristics.

In summary, the goal of Experiment 1 was to test further the hypothesis that feeling-of-knowing judgments are made on the basis of familiarity with the problem statement rather than on partial retrievals of the answer. Previous tests of this hypothesis (e.g., Reder & Ritter, 1992; Schwartz & Metcalfe, 1992) established that familiarity with the correct response was not the source of feeling-of-knowing judgments. However, in these previous tests, it is possible that familiarity with some other (potentially incorrect) response was the source of the feeling-of-knowing judgments. To rule out this alternative explanation, we included in Experiment 1 problem trials in which a problem was presented but no answers (correct or incorrect) were associated with the problem. If the

partial-retrieval-of-some-answer hypothesis is correct, then participants should not be influenced by the unanswered problem trials. By contrast, if the problem-familiarity hypothesis is correct, then participants' feeling-of-knowing judgments should be influenced by exposure to these unanswered problem trials.

Results and Discussion

The data from 5 of the 25 participants were not analyzed because 4 of the participants selected retrieved fewer than 30 times out of the 252 trials; the 5th did not finish the experiment. Fewer than 2% of the trials were excluded because of inaccurate voice key measurements, and fewer than 1% were excluded because of very slow responding (i.e., more than 2 s to select a strategy). We first present several global analyses of performance in this task demonstrating the general adaptiveness of the participants' strategy selections and then present more direct analyses of the hypotheses under study.

Operator differences. Table 1 presents summary statistics for the two operators (* and #). There are two salient differences between the operators: The sharp problems had a higher retrieval strategy selection rate and a higher false-

alarm rate. It is likely that this difference was due to a bias to select retrieve for sharp as some participants attempted to play beat the clock for those problems. However, because both operators exhibited similar behavior in all respects other than this simple bias, the data were collapsed over operators for all of the analyses.²

Strategy selection time. Participants were generally able to select a strategy before the 850-ms deadline. The mean strategy selection time was lower than 650 ms for both operators (see Table 1). Fewer than 10% of the strategy selections were late (greater than 850 ms), and the late responses occurred primarily in the beginning of the experiment.

Appropriateness of strategy selections. In addition to selecting a strategy quickly, participants' retrieve-compute decisions can be evaluated as appropriate or not. By appropriate, we mean selecting to retrieve if the correct answer could be generated within the retrieval deadline. Furthermore, if a participant chose to calculate and answered quickly, that too was a poor choice. This selection appropriateness (also referred to as accuracy) was measured both in terms of the Goodman-Kruskal gamma correlation advocated by Nelson (1984, 1986) and in terms of d' (Swets, 1986a, 1986b). Gamma takes on values between -1 and 1 and reflects the probability that any pair of items has the same ordering in knowing (e.g., one answer is fast and correct, and the other is slow or incorrect) as it has in feeling of knowing (e.g., one involves a retrieve strategy selection, and the other involves a calculate strategy selection): When the probability is 0, γ is -1 ; when the probability is .5, γ is 0; and when the probability is 1, γ is 1. Both measures were computed for each participant and then averaged.

For both measures, hits were defined as trials in which participants chose to retrieve and correctly answered within 1.4 s. Misses were those trials in which participants chose to calculate but answered correctly within 1.4 s. False alarms were trials in which participants chose to retrieve when they could not answer correctly within 1.4 s. Correct rejections were trials in which participants chose to calculate and either took longer than 1.4 s to answer or answered incorrectly.

Participants' accuracy was generally quite high. For multiplication problems, participants had a mean gamma of .82 and d' of 1.95. For sharp problems, participants had a mean gamma of .87, and d' of 2.07. These levels of accuracy are considered quite good, suggesting that whatever mecha-

Table 1
Participant Means and Standard Errors for All Problems
in Experiment 1

Measure	Multiplication		Sharp	
	Calcula- tion	Retrieval	Calcula- tion	Retrieval
Strategy selected (%)				
<i>M</i>	66.2	33.8	61.1	38.9
<i>SE</i>	4.7	4.7	5.1	5.1
Strategy selection time (ms)				
<i>M</i>	647	607	645	596
<i>SE</i>	33	41	21	29
Late to choose strategy (%)				
<i>M</i>	8.6	11.4	11.9	10.9
<i>SE</i>	1.8	4.7	2.0	3.3
Correct answer times (ms)				
<i>M</i>	7,787	1,415	7,235	1,376
<i>SE</i>	523	389	547	242
Correct answer (%)				
<i>M</i>	90.1	67.6	86.7	65.1
<i>SE</i>	1.5	5.3	1.8	4.6
Incorrect choice of retrieval (% false alarms)				
<i>M</i>		20.4		27.7
<i>SE</i>		3.5		4.4
Both strategy choices				
Feeling of knowing and knowing				
<i>M</i>		.82		.87
<i>SE</i>		.10		.04
<i>d'</i>				
<i>M</i>		1.95		2.07
<i>SE</i>		0.24		0.21

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

² Because of a misprint in Reder and Ritter (1992), Experiment 1 was first conducted with an incorrect version of sharp, which was defined as $[(a * c) + (b * d)] * 3$ modulo 100 for a problem $ab \# cd$ (i.e., multiplication and addition were reversed). As a result, the first 6 participants were given the incorrect version of sharp. To assess whether the participants' behavior with the different versions of sharp differed at all, analyses were conducted on percentage of retrieval decisions, decision times, percentage of late decision times, retrieval answer times, compute answer times, percentage of correct responses, percentage of hits, percentage of false alarms, and d' . Because no differences were found on any of the measures ($ps > .25$), results were collapsed across the two versions.

nism underlies participants' strategy selections, it is overall quite adept. Was this generally effective mechanism fooled by the unusual infrequently answered problems? Unfortunately, it is not possible to conduct d' analyses separately for the two problem types as the participants rarely knew the answer to the infrequently answered problems (participant mean of 9.6% vs. 25.9% of regular problems), and the infrequently answered problems occurred on only one in four trials. Analyses focusing on how participants chose to retrieve or compute for the infrequently answered problems are presented in the next section.

An important component of the preceding analysis, and of the design of the experiment, is the assumption that participants could not calculate the answer in the time allotted to them when they selected to retrieve. If the participants could calculate in that short amount of time, then the initial judgment that participants were making would not be between whether to retrieve or calculate, but rather how much time to assign the calculation process. Although it is intuitive that the fast answering times (e.g., less than 2 s) were due to retrieval, it is conceivable that there were some very rapid computations. Conversely, it could be that some of the slow answers assumed to involve the calculation strategy actually involved retrieval. However, there are several types of evidence to support the view that fast answering times were due to retrieval and that slow times were due to computation. First, participants used paper and pencil to perform their calculations. There was no time to calculate using paper and pencil when it was a nominal retrieval trial. Second, one should not have seen so much use of paper and pencil on the calculate trials if they were just slow retrievals. Finally, the notion that the early trials could be slow retrieval violates common sense: How could they retrieve the answer if they did not know it? Thus, there are several reasons to suspect that participants could not calculate within the retrieval time deadline, that fast answer times actually involved retrieval, and that slow answer times actually involved calculation. In other words, the decision that participants made really was a decision about whether to retrieve or calculate. Now, we turn to analyses of the processes that underlie these strategy choices.

The effect of practice on knowing and feeling of knowing. Do both knowing the answer and feeling of knowing change at the same rate? Do they depend on the same variables? If so, feeling of knowing may be based on an early read of the answer. As one learns to associate an answer with its question (or problem statement), the response time to give that answer will decrease. Thus, correct answer time is a measure of the degree of knowing. Likewise, the probability of selecting the retrieval strategy is a measure of feeling of knowing.

To determine which variables predict feeling of knowing, we conducted a logistic regression predicting the probability of selecting retrieval. The variables under consideration as predictors of knowing and feeling of knowing were as follows: (a) *attempted solutions*, the number of times participants were asked to give the answer to a problem, either by computing or by retrieving an answer; (b) *estimations*, the number of times participants estimated whether

they would retrieve or compute, whether they were then actually allowed to retrieve or compute an answer; (c) *total study time*, the total amount of time spent studying that problem's answer³; and (d) *last study time*, the amount of time spent studying that problem's answer the previous time that it appeared. Estimations and attempted solutions were the primary variables under consideration. Total study time and last study time were used as an attempt to equate problems for study time. However, all of the analyses produced qualitatively similar results when total study time and last study time were not entered into the regression equations.

When the entire data set was considered, both estimations and attempted solutions were strong independent predictors of retrieval selections (see Table 2). This result argues against the early-read hypothesis because estimations should not have independent predictive power over attempted solutions under that hypothesis (because they do not themselves lead to any answer being associated with the problem). The independent contribution of attempted solutions over the contribution of estimations neither supports nor contradicts either of the alternative accounts: Attempted solutions may have increased familiarity because of greater depth of processing on normal problems than on infrequently answered problems.

To test this hypothesis, we analyzed whether total exposure (the sum of total decision time, total answer time, and total study time for each problem) could account for the differences between the effects of estimations and attempted solutions. Figure 2 displays the percentage use of the retrieval strategy as a function of the total exposure to the problem. As can be seen in the figure, when we controlled for exposure time to the problem, there are no differences between normal and infrequently answered problems. That is, there appears to be no added effect of actually computing the answer on the probability of selecting retrieve beyond the added exposure to the problem.

An analysis of covariance with Problem Type (normal–infrequently answered) \times Total Exposure Time conducted on the strategy selections confirmed the results suggested by the graph: The effect of total exposure was highly significant, $F(1, 5037) = 459.3, p < .0001$, whereas the effect of problem type was not significant, $F(1, 5037) < 1$.

There is another reason for why the logistic regressions presented in Table 2 might have underrepresented the effect of estimations on feeling of knowing. Over the entire data set, estimations and attempted solutions were highly correlated because they will only differ for the small set of infrequently answered problems. Thus, the logistic regression may not have accurately separated the independent contributions of estimations and attempted solutions. To address this issue, we conducted a separate logistic regression by using data from only the infrequently answered problems, for which participants evaluated the problems on two of every seven trials (see Table 2). For these problems,

³ Note that total study time is not redundant with the attempted solutions and estimations variables because participants could vary the amount of time they spent studying each problem.

Table 2
Results of Logistic Regressions Predicting Strategy Selection for Choice-on-Time Problems in Experiment 1

Variable	Range	Regression coefficient	SE	Z ^a
All problems				
Intercept constant		-0.708	0.156	4.53*
Participant's own coded variable	0.015-3.657			0.071-9.15*
Last time studied (ms)		-0.00002	0.000007	2.95*
Total time studied (ms)		0.000001	0.000002	0.736
Number of estimations	0-27	0.051	0.011	4.66*
Number of attempted solutions	0-27	0.093	0.011	8.67*
Infrequently answered problems only				
Intercept constant		-0.266	0.312	0.855
Participant's own coded variable	0.668-4.777			1.48-4.51*
Last time studied (ms)		0.000012	0.000021	0.058
Total time studied (ms)		0.000013	0.000043	2.51*
Number of estimations	0-27	0.128	0.033	3.83*
Number of attempted solutions	0-8	-0.281	0.108	2.61*

^aComputed as (coefficient/SE_{coefficient}) in the regression.

* $p < .01$.

estimations (the number of times participants were exposed to the problem) remained a strong independent predictor, whereas attempted solutions (the number of times participants were exposed to the answer) correlated with feeling of knowing in the wrong direction (i.e., more attempted solutions lead to fewer retrievals).

Although these results provide evidence against the early-read hypothesis and support the familiarity hypothesis, there is one potential confound that may rejuvenate the early-read hypothesis: The estimation process may have influenced the participants' ability to retrieve the answer. To assess this problem, we conducted a linear regression on the answer times by using attempted solutions, estimations, total study time, and last study time as predictor variables. We found that estimations had no independent predictiveness for answer times, $t(1426) = 0.01$, $p > .99$, above the predictiveness of attempted solutions, $t(1426) = -2.10$, $p < .05$, and last study time, $t(1426) = 2.20$, $p < .03$. If estimations were based on an early read of the answer, then one would have expected estimations to predict answer time.

In summary, the results of this experiment support the familiarity hypothesis and argue against the early-read hypothesis. It cannot be the case that participants were basing their decisions on an early read of some answer (either correct or incorrect) because they were influenced by exposures to problems in which no answer was strengthened.

In the SAC model section, we present a model of the retrieve-compute strategy selection process and fit it to the data generated by Reder and Ritter (1992) and our Experiment 1. This modeling section serves three important functions. First, it provides a mechanistic account of a strategy selection process on the basis of problem familiarity. Second, it demonstrates that a familiarity-based account can provide a good quantitative fit to strategy selection data (including making quantitative predictions about the decay of feeling of knowing with time that we test in Experiment 2). Finally, the model serves to address the

computational conundrum associated with familiarity-based accounts (i.e., the question of why would one try to compute a feeling of knowing rather than simply trying to retrieve immediately).

The SAC Model

Overview

Reder's model is based on what could be viewed as a generic semantic network model of memory (see also Kamas & Reder, 1994; Reder & Gordon, 1996; Reder & Schunn, 1996). The model is called SAC, which stands for

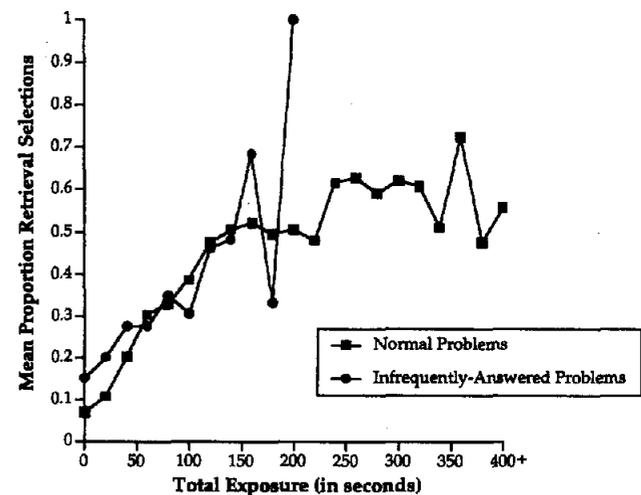


Figure 2. For regular and infrequently answered problems for Experiment 1, the mean percentage of retrieval strategy selections as a function of total exposure to the problem (in seconds). It should be noted that the last three points for the infrequently answered problems are unstable because they have a mean of 19 observations/point, whereas the preceding points have a mean of 150 observations/point.

source of activation confusion. The representation used by the SAC model consists of interassociated nodes representing concepts that vary in long-term strength. In this article, we apply the SAC model to the arithmetic experiment of Reder and Ritter (1992) and to our Experiments 1 and 2. For these simulations, nodes represent numbers, operators, and whole problems. The nodes representing whole problems connect the operands and operators to the answers. Nodes representing numbers may serve as operand nodes for some problems and answer nodes for other problems (e.g., 31 is an operand in the problem $23 * 31$ and is also the answer to $14 + 17$). See Figure 3 for an illustration.

Each node has a base-level or long-term strength. The strength of a node represents the history of exposure to that concept, with more exposure producing greater strengthening. Nodes that represent arithmetic problems such as $27 * 34$ would start out weak at the beginning of the experiment, as these problems were initially unfamiliar to the participants. By contrast, nodes for familiar problems such as $4 * 7$ or $12 * 12$ would be strong, even at the beginning of the experiment. However, the experiments did not use problems that were likely to have preexperimental familiarity, and the simulations presented here assume that all problem nodes are created for the first time during the experiment.

Strength can also be thought of as the baseline or resting activation level of a node. Increases and decreases in this baseline strength change according to a power function:

$$B = c \sum t_i^{-d}, \quad (1)$$

where B is the base-level activation, c and d are constants, and t_i is the time since the i th presentation. This function captures both power-law decay of memories with time and power-law learning of memories with practice.⁴ The central feature of power-law decay is that memories decay quickly initially and then much more slowly at increasing delays. Similarly, the central feature of power-law learning is that first exposures to an item contribute more than do subse-

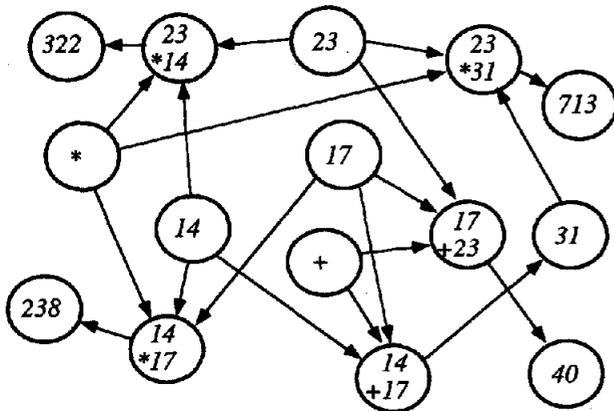


Figure 3. An example semantic network representing problem components, problems, and answers nodes. Note that only a small set of the links emanating from the operand and operator nodes is shown.

quent exposures. That is, the incremental contributions of each new exposure decreases with increasing numbers of exposures.

In addition to the base or resting level of activation of a node, there is also the *current activation* level of a node. The current level of a node will be higher than its baseline whenever it receives stimulation from the environment (i.e., when the concept is mentioned or perceived, or when the concept receives activation from other nodes). Although baseline strength decays according to a power function (i.e., first quickly and then slowly), current activation decays rapidly and exponentially toward the base level. Let A represent the current level of activation and B represent the base level of activation. Then, the decrease in current activation will be

$$\Delta A = -\rho(A - B), \quad (2)$$

such that, after each trial, the current activation will decrease for every node by the proportion ρ times that node's current distance from its base-level activation. To present a concrete example, suppose after a trial, a node's base-level activation was 20 and its current activation was 60. Further suppose that ρ is set to 0.8—the actual value used in all of our simulations. Then, after just one trial, the current activation would drop to 28 [i.e., $60 - .8 * (60 - 20)$], ignoring for the moment the small power-law decay in the base-level activation. After three trials, the current activation would have dropped to 20.3, not significantly different from the resting activation of 20. Thus, current activation drops quite rapidly and only has noticeable effects on the trial on which it became activated, and perhaps the trial immediately thereafter.

Activation spreads between nodes via links. Links connect nodes that are associated through conceptual relations. For example, links connect nodes that represent the components of a problem—operands and operators—to the node that represents the entire problem. Links also connect the nodes representing the entire problems to the nodes representing the answers. These links will vary in strength depending on how often the two concepts have been thought of concurrently. Strength of links also depends on the delay between exposures. Specifically, link strength is determined by a power function given by

$$S_{s,r} = \sum t_i^{-d_L}, \quad (3)$$

where $S_{s,r}$ is the strength of the link from the node s to node r , t_i is the time since the i th coexposure, and d_L is the decay constant for links.

The current activation level of a node can rise from environmental stimulation or from associated nodes that send activation to it. The amount of activation that is sent depends on the activation level of the source (sending) node and on the strength of the link from the source node to the receiving node, relative to the strength of all other links

⁴ See Anderson and Schooler (1991) for a discussion of the evidence for this function in learning and retention phenomena.

emanating from the same source node. The change in activation of some node r is computed by summing the spread of activation from all source nodes s directly connected to node r according to the equation

$$\Delta A_r = \sum (A_s * S_{s,r} / \sum S_{s,i}), \quad (4)$$

where ΔA_r is the change in activation of the receiving node r , A_s is the activation of each source node s , $S_{s,r}$ is strength of the link between nodes s and r , and $\sum S_{s,i}$ is sum of the strengths of all links emanating from node s . The effect of the ratio $S_{s,r} / \sum S_{s,i}$ is to limit the total spread from a node s to all connected nodes to be equal to the node s 's current activation A_s . For example, if a node had three connections emanating from it with link strengths of one, two, and three, then the activation spread along those links would be, respectively, $1/6$, $1/3$ (i.e., $2/6$), and $1/2$ (i.e., $3/6$) of the node's current activation level. Equation 4 is very similar to ones given by Anderson (1976, 1983, 1993) that account for data in fan effect paradigms (e.g., Anderson, 1974). Fan effect experiments have found that the greater the number of competing facts involving a particular object, the slower participants are to accept or reject statements about that object. We offer an analogous explanation that also involves competition for the limited activation from a source node. It is important to note that the absolute magnitude of the link strength is irrelevant—only the strength relative to the total strength of other links emanating from the same node matters. By contrast, the absolute magnitude of activation level of the sending and receiving node does matter.

In this spreading-activation model, feeling-of-knowing judgments are based on the activation level of the node representing the problem. In essence, it is assumed that feeling of knowing monitors intersection of activation from two source nodes. Specifically, when two terms in a problem send out activation to associated concepts, and an intersection of activation is detected by bringing an intermediate node over threshold, a person will have a feeling-of-knowing response (cf. Doshier & Rosedale, 1989, 1990; Glucksberg & McCloskey, 1981; Ratcliff & McKoon, 1988; Reder, 1979, 1987, 1988, for related treatments of intersection of activation).

In our current simulations, we assume that when a problem is presented, all of the nodes representing the components are activated. For example, in the problem $23 * 14$, the nodes representing 23 , $*$, and 14 are all activated. Then, activation spreads from the component nodes to all the connected problem nodes (see Figure 3). In the $23 * 14$ example, activation spreads to all the problem nodes involving 23 (e.g., $23 * 14$, $23 + 17$), $*$ (e.g., $23 * 31$, $14 * 17$, $23 * 14$), and 14 (e.g., $14 + 17$, $14 * 17$, $23 * 14$). Problem nodes connected to several of the components receive the greatest amount of activation (e.g., $23 * 14$). The extent of activation that accumulates at the problem node affects the likelihood of selecting retrieve as the strategy of choice. In a similar fashion, activation spreads from problem nodes to answer nodes. This is how answers are retrieved. Relations of connectivity define the objects, but a given node can be

both an answer and an operand (e.g., 31 as the answer to $14 + 17$ and as an operand in $23 * 31$).

Because activation that spreads to a node is added to the base activation, the selection of which problem node will have the highest final activation will also depend on the relative base-level activations. The current activation level of the most (currently) active problem node is used to determine feeling of knowing. On the basis of the feeling of knowing, a decision is then made to retrieve or compute. That is, if the problem node has a relatively high activation level, then retrieval will most likely be selected; if the problem node has a relatively low activation level, then computation will most likely be selected.

This model unravels the computational conundrum underlying the use of feeling-of-knowing judgments in strategy choice—feeling of knowing is automatically represented in the parsing and representing of the problem. Thus, in this scheme, feeling of knowing is a natural precursor to the retrieval process.

Model Details

In addition to predicting feeling-of-knowing decisions (i.e., decisions between retrieval and computation), this model can also predict which answers are retrieved from memory, and the speed with which the answers are retrieved. In this article, however, we focus on the feeling-of-knowing, or retrieve-compute, decisions. As input, the computer simulation is given the same problems presented to each participant. Because each participant received a different set of problems in random order, a separate simulation was conducted for each participant. This precise yoking of the simulation to participants was important because on a given trial the expected activation level for a problem would vary depending on the exact sequence of trials: For any participant on a given trial, the number of links, the current activation, and the strengths would be different from any other participant's values. The simulation output is a probability of selecting to retrieve on each trial. We now step through the process by which that probability is determined.

At the start of the experiment, the representation of memory for the simulation is identical regardless of the experimental stimuli to be seen. Nodes for the operands are assumed to already exist; whereas nodes for the problems are assumed not to exist (i.e., the problems are novel). For simplicity, the initial base-level strengths of the operand nodes (numbers used as problems) and operator nodes are set to a constant amount, the amount being irrelevant to the simulations of the retrieval process. When problems are seen for the first time, a problem node is created, as are the links from the component operand and operator nodes to the novel problem node. The initial base-level strengths of the problem nodes and of the links are simply determined by the equations determining power-law growth and decay—the computation of initial strength values requires no extra parameters.

On each trial, all of the nodes representing the problem components are activated to the same constant amount, again for simplicity. It is assumed that a basic perceptual

process activates these nodes and that all of the problem components (e.g., the operators and the operands) used in these experiments were familiar entities. For example, when the problem $23 * 14$ is presented, the three nodes representing 23, *, and 14 are activated. Activation then spreads along the links emanating from nodes representing each of the problem components to nodes representing the complete problems themselves. Activation only spreads to directly connected nodes at this point and is not yet carried forward beyond the first layer of receiving nodes. Again, the amount that is spread is a function of that specific participant's problem presentation history.

Once the activation has spread across these links, activation of the problem nodes can be used to make a strategy selection between retrieve and calculate (feeling of knowing). The activation value of the most active node is used. It is assumed that this decision follows a normally distributed function of activation. Rather than making a binary choice, the simulation predicts a probability of choosing retrieve on the basis of this activation value. This means that if the activation value of the most active node is low, the probability of selecting retrieve is very low; conversely, when the activation value of a node is very high, the probability of selecting retrieval is high, but not necessarily unity. This probability of choosing retrieve is calculated by assuming a normal distribution of activation values with a fixed variance and activation threshold for selecting retrieve. This probability is reflected in the formula

$$P = N[(A - T)/\sigma], \quad (5)$$

where A is the activation of the most-active problem node, T is the participant's threshold, σ is the standard deviation, and $N[x]$ is the area under the normal curve to the left of x for a normal curve with mean equaling zero, and standard deviation equaling one.

After each trial, all of the node strengths and activations are updated by using Equations 1 and 2. Link strengths are also updated for each link, following the same kind of

power-law function used to determine changes to base-level activation—all of the links connecting the problem component nodes to the problem node in the just-presented problem are strengthened; whereas all other links in the network are weakened (using Equation 3 for both strengthening and weakening). It is at this point that if a new problem has been presented for the first time, then a new node representing that problem is created, and links are created connecting the component nodes to the problem node. As with the initial strength values of the nodes, the initial strength values of the links are determined by the growth and decay equation—no extra parameters are required. This process of updating nodes and links is identical whether the participant actually selected to retrieve or calculate, or whether a correct or incorrect answer was given (because the participants were always given the correct answer to study). Although it is likely that the amount of time spent studying the answer will influence the strengthening, as a simplifying assumption a common increment amount independent of study time is assumed.

The simulation just described involves seven parameters that are listed in Table 3. Two of these parameters are related to the initialization and decay of current activation. First, the input-activation parameter, arbitrarily set to 50 for all of the simulations, determines the current activation setting of the nodes representing the problem components (but not problems nodes) when the problem is presented. Second, the fast-decay parameter, ρ , is the exponential decay constant at which the current activation of all nodes decays. For simplicity, the unit of decay is trials rather than time. Preliminary analyses indicated that a value of 0.8 for the decay parameter gave the best fit, and thus this parameter value was used for all simulations.

Two other parameters of these seven are necessary for changing the base activations. These are the two parameters in the power-law Equation 1 determining base activations, c and d . They were set to 5 and 0.175, respectively, for all of the simulations. Thus, the initial strength value of problem node after its creation was 5, and it decayed with time and

Table 3
SAC Model Parameter Descriptions and Values and the Model Equations

Parameter name	Function	Value
Input activation	Input current activation for component nodes	50
ρ	Exponential decay constant for current activation	0.8
c	Power-law growth constant for base-level activation	5
d	Power-law decay constant for base-level activation	0.175
d_L	Power-law decay constant for link strength	0.12
T	Retrieve-compute decision threshold	30–200
σ	Retrieve-compute decision standard deviation	45
Never retrieve	Does participant decide to never retrieve for one of the operators?	True–False
Equation	Description	
(1) $B = c\sum t_i^{-d}$	Base-level activation as a function of delay and repetitions	
(2) $\Delta A = -\rho(A - B)$	Change in current activation from one trial to the next	
(3) $S_{s,r} = \sum t_i^{-d_L}$	Link strength as a function of delay and repetitions	
(4) $\Delta A_r = \sum (A_s * S_{s,r} / \sum S_{s,r})$	Change in receiver's current strength due to activation spread	
(5) $P = N[(A - T)/\sigma]$	Probability of selecting retrieve as a function of current activation	

Note. SAC = Source of activation confusion.

grew with repeated presentations from there. As with fast decay, trials were used as the unit rather than time for simplicity.

Only one parameter is used in the computation of link strength. Link strength is updated using the same kind of power-law function used to determine base activation. However, the c constant is not used because the absolute magnitude of the links is irrelevant (because of the effect of fan in Equation 4). The d constant, labeled d_1 to differentiate from the d constant used in the determination of base-level node strength, was set to 0.12.

To convert these activation values to probabilities of selecting retrieve (vs. compute), two further parameters are necessary. Recall that we assumed this decision follows a normally distributed function of activation. Correspondingly, there are two parameters used to determine the shape of this normal function: the threshold which is the center of this distribution and the standard deviation. A single value for the standard deviation parameter was used for all simulations.

However, in contrast to the single standard deviation, we assumed that participants vary in their thresholds for choosing between retrieve and compute. That is, some participants were conservative and had high thresholds; other participants were optimistic and had lower thresholds. This threshold value reflects the participant's overall base rate of selecting retrieve. Note that because this value is the center of the normal distribution, the probability of selecting retrieve when the activation value equals the threshold was .5. A value between 30 and 200 was selected for each participant to maximize fit to their data. This wide range of possible values mirrored the large between-participants variance that was found within each of the experiments in the retrieval selection rates. Although the participants might have differed on other dimensions as well, there were no other obvious differences (with the exception of the one mentioned below), and so, for parsimony's sake, the other six parameters were held constant across participants.

There is one final component of the SAC model that required an additional parameter. This eighth parameter was only used for simulating some of the participants. The parameter was simply a binary value by participant reflecting whether the participant had a predilection not to choose retrieve for a particular operator. This parameter was added because we found that some participants had a strong aversion to choosing retrieve for a particular operator. For example, a few participants never chose retrieve for problems involving the operator sharp (a novel operator that involved a combination of addition and multiplication). Perhaps they did not want to memorize problems that involved a fake operator. A few other participants were found to never retrieve for multiply, although they chose retrieve for sharp problems. These participants may have been bothered by the modular arithmetic that was used in some of the experiments and did not want to memorize the wrong answers to multiply problems. Whatever participants' reasons for choosing to never retrieve for an operator, we found that the eighth parameter was useful for simulating these participants—those that seemed to use a metarule for

making their decisions, in which they refused to retrieve for one of the operators. To model these participants, the probability of selecting retrieve on that operator was set to zero. For those participants, the probability of selecting retrieve for the other (nonmetarule) operator was simply determined as for the regular participants—by the equations given in the SAC model. A simple 5% cutoff was used to select which participants to model with this never-retrieve rule: Participants had to have selected retrieve for less than 5% of the trials with a particular operator. The evidence for the use of this rule is presented with the simulations.

It should be noted that this rule was invoked only eight times out of the 58 participants modeled. We felt it was better to use this metarule than to assign separate thresholds for problems of each operator type. Not only would this give us too many degrees of freedom, it was hardly necessary: Except for these few participants using this metarule, the correlation between the rate of selecting retrieval for problems involving each operator type was quite high across participants. Finally, it is important to note that although we believe that some participants actually used this metarule, this feature of the simulation is not necessary to fit the data. Therefore, the fits to data without the use of this feature are also presented.

In summary, there are eight parameters for the simulations, six of which were held constant for all simulations. Table 3 presents a summary of these parameters, as well as the five equations underlying the SAC model.

To compare SAC's predictions to participants' actual retrieve-compute decisions, we used an aggregation procedure developed by Anderson (1990). For each trial, for each participant, the model produced a probability of choosing retrieve on the basis of the calculated activation values resulting from the trial history for that participant. That is, the probability reflected the model's experience with the exact same problems given to the participant. This probability was also based on the particular participant's threshold. Because participants made binary decisions on each trial and the simulation produced probabilities, it was necessary to aggregate trials. That is, all trials for a given participant for which the simulation predicted that the probability of selecting retrieve would fall between 0% and 10% were grouped together; all trials in which the probability fell between 10% and 20% were grouped together and so on. Next, we tabulated the actual proportion of retrieval strategy selections that were made by that participant for the exact same trials in each probability range. This was done for all probability ranges. Note that each participant contributes data points to each (or at least many) of the ranges. The fit of the model was tested by plotting mean actual proportion of retrieval strategy selections against mean expected proportion of retrieval strategy selections. A perfect fit would be a straight line with a slope of one and a y-intercept of zero (i.e., predicted = actual). On each graph, we plot this desired line to show where the fitted points should actually lie.

Rather than plot the full scatter plot of each participant's value in each probability range, which often contains too many points from which to abstract the central tendency

accurately, we plot the mean participant value (i.e., mean of participant means) within each range. To present an estimate of the participant variance, we also plot standard error bars. Furthermore, we present the r^2 value between predicted and actual values on the basis of the full scatter plot, not the mean responses across participants. This value presents a good estimate of the amount of variance that the model accounts for at the individual participant level, a fine-grained level of detail not typically presented in tests of computational or mathematical models. To assess whether there are any systematic biases in the model's predictions, we also present the slope and y -intercept of the best fitting regression line.

Because there are a differing number of observations that contribute to each participant's value, and they vary across the probability ranges, we also plot the number of observations that fell within each probability range. The number of observations often varied widely across the probability ranges because values were much more likely to fall into the lower probability ranges. Therefore, we felt a logarithmic scale was most appropriate.

Because the number of participants and data points per participant varied for the various experiments and analyses, it was necessary to vary the size (and hence number) of the probability ranges. If the ranges were made too small, then participants would contribute too few observations to each range, and the participant estimates would contain too much instability because of random noise. If the ranges were made too large, then there would be very few points plotted, and too little information would be given about the quality of the fit. Compromise values were selected for each analysis by using the following rule of thumb: The ranges were made sufficiently large such that each participant contributed at least five data points to most of the ranges, thereby ensuring stable proportions. Note that for a given analysis, all of the probability ranges are of the same size: $1/n$, where n is the number of ranges used. In the few cases in which a participant contributed only two or fewer data observations to a probability range, that participant was considered not to have contributed anything to that probability range. This procedure was necessary because proportion estimates are highly unstable for n s of one and two.

We used the values produced by this aggregation procedure to derive the best fitting parameters. The fit of the model to the data was defined as the sum of the squared error between the model's predicted retrieval rate for each participant in each range and each participant's actual retrieval rate in each range. The full, exhaustive combinatorial space of possible parameters was not searched. Instead, we iteratively tried a range of values on each parameter dimension, selecting the value on each dimension producing lowest sum squared error. For the six parameters held constant across the simulations of different data sets, the parameters were determined once for the first data set and were then held constant across all other data sets. Although we did not conduct an exhaustive search of the parameter space, we did find that changing any of the parameter values by more than 20% did alter the fit of the SAC model in a way such that selecting new values for the other parameters could not

completely compensate. In other words, all of the parameters did contribute to the fit of the SAC model.

Here is a brief summary of the process of fitting the SAC model to one participant's data. The model begins with only operator and operand nodes. When problems are seen for the first time—in the order the participant was given the problems—problem nodes and the links connecting them to the component (operand and operator) nodes are created. For a given trial, the model activates the operator and operands from the problem being presented. Activation spreads to all directly connected problem nodes. The activation of the problem node with the highest activation is used to make the feeling-of-knowing judgment. This most-activated problem node is usually the one that represents that problem; however, it need not be. For example, if the problem node for the currently presented problem does not exist, then, of the other problem nodes that have partially been activated, the one with the highest activation value is selected.⁵ If the current participant was determined to be using the never-retrieve metarule for the current operator, then the probability of selecting retrieve is simply zero. Either way, the SAC model generates a predicted probability of selecting retrieve. This is the prediction that is compared against the actual response by using the aggregation procedure described earlier. The network is then updated as follows: (a) new problem nodes and links are created if necessary, (b) baseline activations are updated for all nodes by using Equation 1, (c) current activations are updated for all nodes by using Equation 2, and (d) link strengths are updated for all links by using Equation 3. Because we were not modeling the answers that participants gave, the SAC model was not influenced by the answers that participants gave or by the time to give them. This process is then repeated for each trial for a participant, and the whole process is repeated for each participant. At the end, the mean proportion of actual responses is calculated for each participant in each predicted probability range.

Simulation of Reder and Ritter (1992)

As a first test of the SAC model, we selected the strategy choice data from Reder and Ritter (1992). We focused on data from their Experiment 2. Reder and Ritter's Experiment 1 involved addition and multiplication facts. For the addition problems, many participants always selected retrieve and tried to quickly compute the answer, as there was a heavy incentive to select retrieve if the answer could be given on time. To remove this problem, Reder and Ritter's Experiment 2 replaced addition with a new operator (the sharp operator used in our Experiment 1) that participants could not execute as quickly. Because Experiment 2 did not have the methodological problem just mentioned, it is this experiment that we used.

Reder and Ritter's (1992) Experiment 2 used the same methodology as our Experiment 1, with a few exceptions, which are noted below. The primary exception was that they

⁵ On the very first trial, there are no problem nodes, so an activation value of zero is used for the feeling-of-knowing judgment.

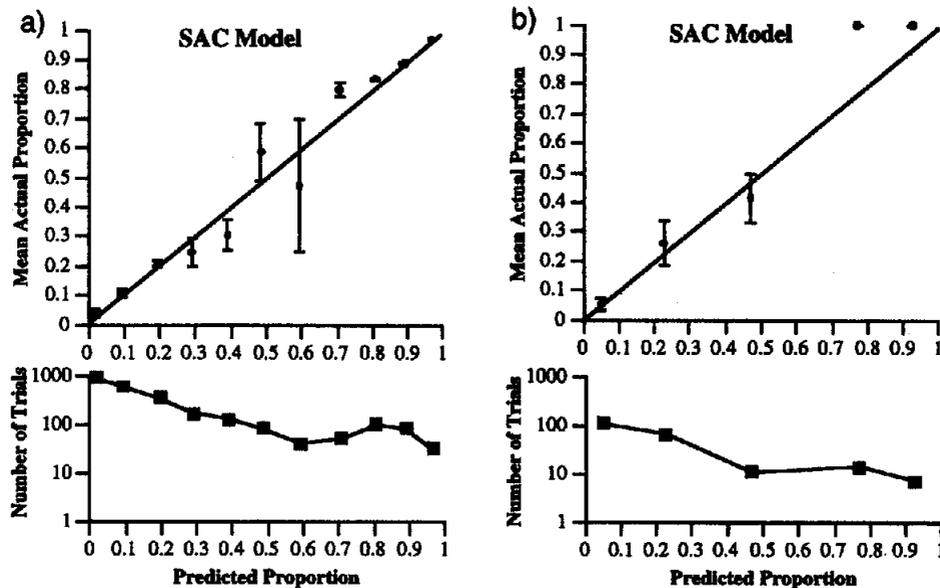


Figure 4. For (a) all problems and (b) the operator-switch problems only, in Reder and Ritter (1992), mean actual proportion of retrieval strategy selections (and standard error) as a function of grouped predicted proportions of retrieval strategy selections, along with the number of trials in each grouping plotted in logarithmic scale. Note that the line drawn in the graph is the desired line actual = predicted, not the best fitting regression line. SAC = source of activation confusion.

did not have infrequently answered problems—their participants had to provide answers to all problems, as in our Experiment 2. Instead, Reder and Ritter used operator-switch problems (a training problem with the operator switched to be the other operator). In their experiment, participants were asked to solve 18 arithmetic problems repeatedly. Half of the problems involved multiplication, and the other half involved the sharp. In addition to 200 training trials with these 18 problems, operator-switch problems were interspersed among them.

Reder and Ritter (1992) found that participants' retrieve-compute decisions were predicted by the frequency with which the two operands appeared together. Furthermore, participants were just as likely to select retrieve for the corresponding operator-switch problems (for which they did not know the answers) as for the training problems. Reder and Ritter interpreted these results as supporting the hypothesis that participants were using familiarity with the problem rather than with the answer in deciding whether to retrieve or compute.

To further test this hypothesis, we compared SAC's predictions to participants' actual retrieve-compute decisions, using the aggregation procedure described earlier. The best fitting participant thresholds ranged from 30 to 200, with a mean threshold of 130.8 ($SD = 42.8$). Using these values, SAC fit the data quite well, producing a Pearson's r^2 of .85 (see Figure 4a).⁶ The slope of the best fitting line was not significantly different from one (slope = 0.993), $t(56) = 0.125$, $p > .9$, nor was the intercept significantly different from zero (intercept = -0.001), $t(56) = 0.029$, $p > .9$. In other words, the SAC model accounted for a large percentage of the variance of the participant's strategy selections

even at the individual participant level, and there were no systematic biases in the model's predictions.

A key result of Reder and Ritter (1992) was that participants were as likely to select retrieve for operator-switch problems as for the training problems. The SAC model predicts this effect: Operators are associated with a large number of problems (i.e., they have a large fan out). The activation spread from a node along each link is inversely proportional to the total connection strength of the links emanating from that node. Thus, very little activation spreads from operator nodes to any particular problem node (see Anderson, 1983, for a more detailed discussion of the fan effect). Accordingly, the SAC model predicts that there will be little impact of switching operators on retrieve-compute decisions because the activations of the problem nodes are not significantly affected. Verifying this prediction, the fit of the SAC model to the operator-switch retrieve data was quite good ($r^2 = .82$). Figure 4b presents this fit. Fewer groupings were used in this analysis because there were relatively few operator-switch problems. Again, the slope of the best fitting line was not significantly different from one (slope = 1.17), $t(23) = 1.42$, $p > .15$, nor was the intercept significantly different from zero (intercept = -0.009), $t(23) = 0.22$, $p > .8$.

⁶ Reder and Ritter (1992) found that participants learned "twin" problems (i.e., same top and bottom operands) much more quickly than nontwin problems. This feature was not added to the model, and, consequently, the twin data were not included in the model fits.

An additional, noncentral assumption of the SAC model involved the never-retrieve rule. What is the evidence for this rule? The motivating factor for this rule was that some participants almost never selected retrieve for one of the operators while selecting retrieve quite frequently for the other operator. Because problems with either operator were presented equally frequently and the solutions were equally complex, there was no external, environmental basis for this difference in retrieval selection rates. With the 5% cutoff used for selecting never-retrieve rule users, 4 participants were assumed to be using this rule. Using data from all the participants, the correlation between participant retrieval selection rates for one operator and participant retrieval selection rates for the other operator was .93. When the 4 participants were removed, this correlation rose to .96. This increase in correlation size is noteworthy given that the usual effect of reducing the range of data is to lower correlations. However, the fits of the SAC model did not crucially depend on the application of the never-retrieve rule. In fact, the overall correlation between actual and predicted retrieval rates without the never-retrieve rule ($r^2 = .89$) was slightly higher than with the rule.

Value of Each Parameter

One potential criticism of the SAC model is that it contains many free parameters. This leads to the question: Are all of the parameters necessary, or could simpler models provide an equally good fit? We have already evaluated the necessity of the never-retrieve rule and found that, in this case, it was not necessary for the excellent fits to data that we obtained. Rather than testing the value of the remaining parameters individually, we address this issue more globally by exploring one particular reduced alternative model. This alternative model might be called the *everything-is-in-the-threshold-values* account. Because each participant was given a different threshold value, and there were more participants than probability ranges, it might be argued that the good fits were due to having more free parameters than data points. The obvious response to this criticism is that (a) each participant contributed to most of the probability ranges, and (b) the correlations were based on the individual participant data rather than on aggregations across participants. However, to evaluate this alternative more formally, we created a variant of the SAC model in which the model's predictions for each participant were scrambled. That is, the original model's predictions for each participant were kept, but the pairing with the participant's actual responses was randomly reorganized. For example, rather than having the model's prediction for the first trial paired with the participant's response to the first trial, the model's prediction for the first trial might be paired with the participant's response to the 10th trial, or perhaps the 100th trial. This randomization was done separately for each participant. This method holds constant the distribution of predictions made for each participant and the distribution of responses made by each participant but removes the contingency with the ordering of trials and the types of trials. The logic of this test was as follows: If the original model's predictions were entirely due

to the participant thresholds, then only the participant thresholds should matter, and the rest of the SAC model components based on power-law growth and decay and so forth should have no effect.

In fact, this scrambled model was able to account for 54% of the variance in the individual participants' strategy selections, suggesting that participant thresholds were an important part of the SAC model's good fit to the data. However, this fit is much worse than the 85% of the variance for which the original model can account (see Figure 5a). Furthermore, the scrambled model's best fitting regression line deviates significantly from the desired line: Its slope differed significantly from one (slope = 0.694), $t(56) = 3.56$, $p < .001$, and the intercept differed significantly from zero (intercept = 0.068), $t(56) = 2.04$, $p < .05$. In other words, in addition to the participant thresholds, the rest of the SAC model's machinery was necessary to produce the good fits observed in Figure 4. The values for these other parameters will be held constant in the remaining data fits, providing a strong test of the model given that these parameters do play an important role in the quality of the fits.

Comparison With Other Models

It is worth noting that the SAC model's predictions for this aspect of the data are in direct contrast with the predictions of other models of cognition. For example, Logan's (1988) Instance Theory assumes that there is a race between retrieving the answer and computing the solution, and that the speed with which answers are retrieved is dependent on the number of instances of that answer that are stored in memory. Because no instances of the answer to the operator-switch problems had been stored in memory, Logan's theory would predict that participants would never attempt to retrieve the answer. Under Logan's theory, it cannot be the case that participants were simply not encoding the operator, because problems with the two operators were intermixed throughout the task, and the participants needed to use information about the operator to decide which computing strategy (i.e., algorithm) to use. In other words, if it is assumed that the calculation strategies are executed in parallel with the retrieval process, then the participants could not be ignoring the operators and retrieving the wrong answers, because they had to have been immediately encoding the operators to begin the calculation process.⁷

As another alternative to the SAC model, there is a class of strategy selection models that we call *base-rate models* (e.g., Anderson, 1993; Lovett & Anderson, in press; Siegler & Jenkins, 1988; Siegler & Shipley, 1995). Base-rate models assume that strategies are selected according to the relative proportion of times each strategy has been successful. Such a model could be quite fruitfully applied to our experimental data as it makes some correct qualitative predictions. In

⁷ Of course, under our account, it is not necessarily the case that participants were encoding the operators—the data from our Experiment 1 can address this possibility. It is only Logan's (1988) theory that requires that participants encode the operators.

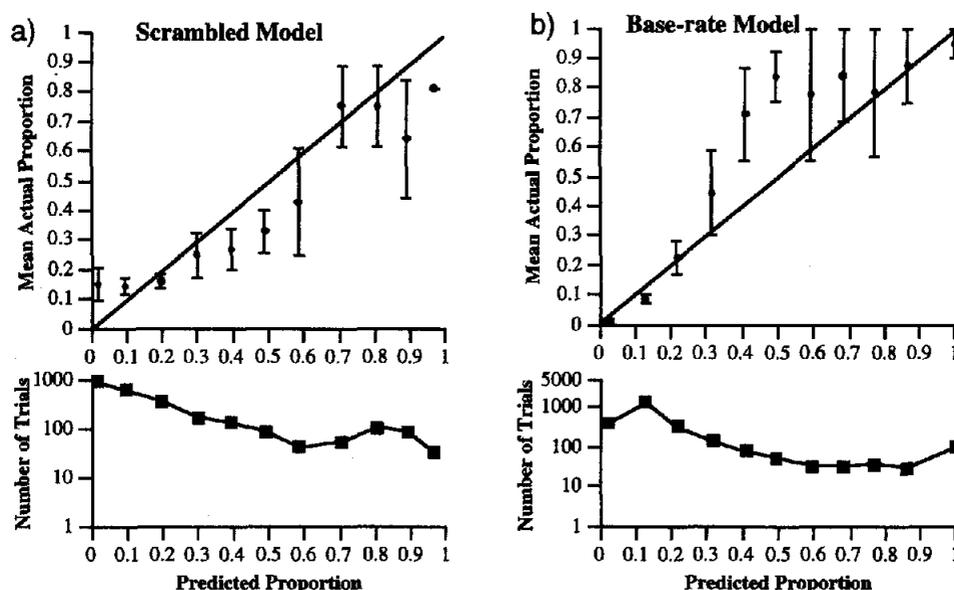


Figure 5. For all problems in Reder and Ritter (1992), using (a) the scrambled model and (b) the base-rate model, mean actual proportion of retrieval strategy selections (and standard error) as a function of group predicted proportions of retrieval strategy selections, along with the number of trials in each grouping.

particular, such a model would predict that participants should initially select to calculate and gradually shift to selecting to retrieve because the experiment was designed such that participants would initially know none of the answers and gradually know an increasingly larger percentage of the answers.

To evaluate whether such a base-rate model could account for as much variance as the SAC model, we tested the following model, using the same evaluation procedure used for the SAC model. We assumed that there was a linear increase over trials in the probability of selecting retrieval because analyses of the data had suggested that there were no significant curvilinear trends over time. Each participant was assigned two parameters: the initial retrieval rate,⁸ and the rate at which retrieval selections increased over time. The best fitting parameter values were used. Despite having many free parameters, the base-rate model was only able to account for 71% of the variance in the individual participant strategy selections (see Figure 5b), significantly lower than the 85% produced by the SAC model. The slope of the best fitting line was not significantly different from one (slope = 1.09), $t(39) = 0.75$, $p > .45$, nor was the intercept significantly different from zero (intercept = 0.032), $t(39) = 0.61$, $p > .5$. Yet, Figure 5b shows that there were serious deviations between the predicted and actual strategy selections rates. Thus, although there may be some use of success base rates in the participants' strategy selections, the SAC model provides a better overall account of the strategy selection data. Furthermore, the base-rate model could not explain why participants would be sensitive to the familiarity of operator-switch problems—it would simply predict that the current base rate would be used no matter what the familiarity of the operator-switch problem.

In summary, the SAC model presents a very good fit to the data from Reder and Ritter (1992). In addition to providing a strong fit to the training data (much stronger than several alternative models), the SAC model also provided a good account for the operator-switch data. However, it might still be argued that the close fits to the data may be due to the high degrees of freedom associated with a model with eight parameters. To provide a stronger test of the SAC model, it would be desirable to fit a different set of data by using the same parameter settings. The data from Experiment 1 provided an opportunity for such a test.

Experiment 1 Simulations

As with the fit to the Reder and Ritter (1992) data, we compared the SAC model's predictions to participants' actual retrieve-compute decisions. Again, the probability of selecting retrieve for each trial was computed, and trials were grouped according to ranges of expected probability. Because our new Experiment 1 produced more data than Reder and Ritter's experiment, a greater number of (smaller) groupings were used. To provide a much stronger test of the SAC model, we set the model's parameters to the same values that were used in the simulation of Reder and Ritter. The only parameters that we did not take from the simulation of Reder and Ritter to use for the simulations of Experiment 1 were the two participant-specific parameters: the participant's threshold, and whether they used the never-retrieve

⁸ This initial retrieval rate is equivalent to a participant-specific threshold. Thus, this alternative model subsumes the individual threshold model and is guaranteed to provide at least as good of a fit to the data as the scrambled model.

rule for an operator. The best fitting participant thresholds ranged from 50 to 155, with a mean threshold of 115.5 ($SD = 22.8$).

As with the Reder and Ritter (1992) data, the SAC model fit our new Experiment 1 data quite well, producing a Pearson's r^2 of .69 (see Figure 6a). The slope of the best fitting line was not significantly different from one (slope = 0.951), $t(254) = 1.22$, $p > .2$, nor was the intercept significantly different from zero (intercept = 0.011), $t(254) = 0.54$, $p > .5$. Thus, the SAC model generalizes very well to other data sets with all but the participant-specific parameters held constant across the data sets.

The primary manipulation of Experiment 1 was the introduction of infrequently answered problems. Because frequency of presentation is the basis of strategy selections in the SAC model, the model predicts that participants' feeling of knowing for the infrequently answered problems should increase just as rapidly as with the normal problems. The SAC model was able to account for participant's behavior on these infrequently answered problems: The slope of the best fitting line was not significantly different from one (slope = 0.936), $t(105) = 0.68$, $p > .45$, nor did the intercept differ significantly from zero (intercept = -0.036), $t(105) = 0.80$, $p > .4$. Although with a little more noise in the predictions, because of the smaller number of observations per point, the SAC model accounted for a near majority of the variance in the participant-level behavior, producing an r^2 value between the actual and predicted proportion of retrieval selections of .49 (see Figure 6b).

The effect of this manipulation serves to differentiate empirically between the SAC model and several other models of cognition. Recall, for example, Logan's (1988)

Instance Theory, which assumes that there is a race between retrieving the answer and computing the solution and that the speed with which answers are retrieved is dependent on the number of stored instances of that answer. Under such an account, participants should not be affected by the trials in which they did not compute an answer, as no answer is then stored with the problem. This incorrect prediction holds true for other theories of strategy selection as well (e.g., Anderson's, 1993, ACT-R; Siegler & Shipley's, 1995, ASCM [adaptive strategy choice model]), which assume that strategy selections are influenced by the strength of the answer and expectations about the relative speed and successfulness of each strategy that derived from previous experience. Because the unanswered trials provide no information about the answer nor about the speed and successfulness of strategies, these other theories would also predict no effect of the unanswered problem trials. Yet, participants were affected by unanswered trials, as predicted by the SAC model. Consequently these other theories cannot account for the main results of Experiment 1.

The data from Experiment 1 provided a better test of the importance of the never-retrieve rule to the SAC model. In this experiment, 3 participants were classified as using this rule under the same criterion used to fit the Reder and Ritter (1992) data. Using data from all participants, the correlation between participant retrieval selection rates for each operator (i.e., how often they selected to retrieve with sharp problems compared with how often they selected to retrieve with multiplication problems) was not statistically significant ($r = .156$, $p > .5$). However, with these 3 participants removed, the correlation was quite high ($r = .786$, $p < .0001$), indicating that all but those 3 participants evaluated the two operators similarly. Thus, Experiment 1

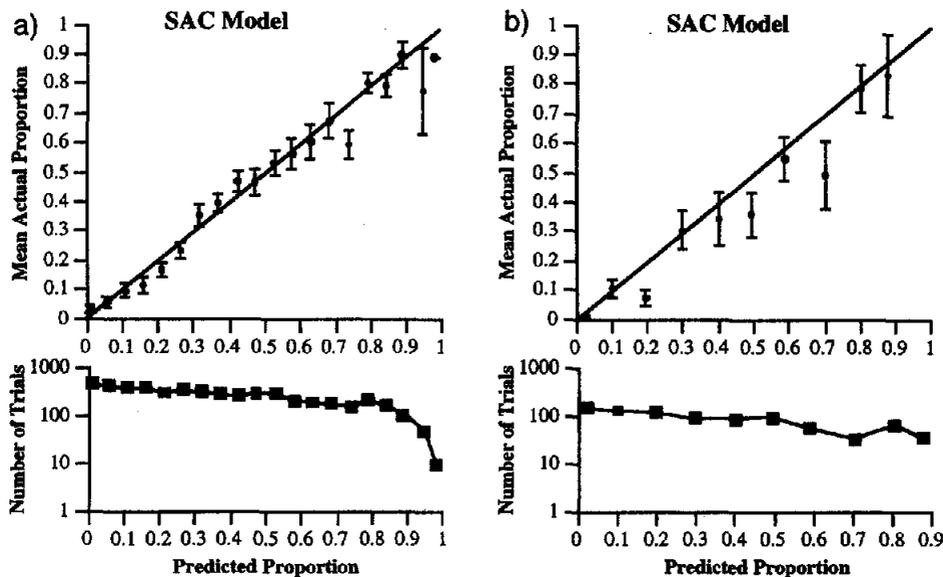


Figure 6. For (a) all problems and (b) the infrequently answered problems, in Experiment 1, mean actual proportion of retrieval strategy selections (and standard error) as a function of group predicted proportions of retrieval strategy selections, along with the number of trials in each grouping. SAC = source of activation confusion.

presented strong evidence for some participants adopting a form of metastrategy similar to the never-retrieve rule.

The fit of the SAC model to the data for Experiment 1 was also more dependent on the use of the never-retrieve rule than with the Reder and Ritter (1992) data. Without the never-retrieve rule, for all problems the SAC model accounted for significantly less variance ($r^2 = .61$), although still at a high level overall. Furthermore, the slope of the best fitting line was significantly less than one (slope = 0.842), $t(256) = 3.76$, $p < .0002$. In other words, without the never-retrieve rule, the model overpredicted the participant's retrieval selection rates.

In summary, with the parameters set to the same values used for the simulations of the Reder and Ritter (1992) data (but still allowing participant-specific parameters), the SAC model produced a very good fit to the data. This high degree of generalization across the fits to both data sets may not be all that surprising given that the paradigm and interface used in Experiment 1 were very similar to those used by Reder and Ritter. In Experiment 2 we sought to provide a stronger test of the SAC model's ability to generalize. In particular, Experiment 2 provided a test of whether the exact size of the long-term effects are accurately predicted by the SAC model, holding constant all of the parameter values (crucially including the learning and decay parameters) from the simulations of Reder and Ritter and our Experiment 1.

In Experiment 2 we also sought to address another remaining problem with past confirmations of the problem-familiarity hypothesis, including the current Experiment 1: Previous experiments on feeling of knowing have used relatively short delays, typically less than 1 hr. As a result, participants may have been making feeling-of-knowing judgments on the basis of problem representations still in working (or active) memory, whose influence overwhelms any influence of an early read of the answer. For more typical, longer term delays, early reads of the answer might become influential on strategy selection. Thus, it is possible that the findings from previous experiments will not generalize to longer delays (i.e., that retrieve-compute strategy decisions at longer delays are based on an early read of the answer rather than on familiarity with the problem statement). Experiment 2 was designed to test this hypothesis.

Experiment 2

The overall design of Experiment 2 was to present participants with a great deal of practice on some arithmetic problems on one day, and then test the participants on related (similar-looking) problems 24 hr later. To make the test situation more realistic, we presented the test problems intermixed with new problems added on the second day. The test problems were operator-switch problems (same operands, new operator) analogous to those used by Reder and Ritter (1992). Thus, if participants select retrieval for these test problems, it is because the problems seem familiar rather than because they know the answer, because they had not before seen the problems with the operators reversed.

Method

Participants. Participants were 29 undergraduate and graduate students at Carnegie Mellon University. They received course credit, money, or both for their participation in the 2-day experiment.

Procedure. A procedure similar to that of Experiment 1 was used. Participants were told that they would be shown a long series of arithmetic problems. After each arithmetic problem was presented on the screen, participants rapidly chose to either calculate or retrieve the answer. They then executed a strategy and gave their answer, after which they were informed of the points earned. Finally, they studied the problem and answer. However, unlike the previous experiments, a new interface was used, the primary differences being that participants entered their responses using the keyboard rather than responding verbally and that participants gave the full answer to the questions rather than the answer modulo 100. The new interface was developed to eliminate voice key errors and to test the generality of the effects to other response modalities. Modulo arithmetic was removed to establish the previous results using a more realistic version of multiplication.

Each trial began when the participant pressed the space bar. The problem was displayed with the words "Retrieve or Calculate" above the problem. The participant then chose a strategy by pressing one of two keys. The zero and period keys were selected as they were on the numeric keypad that participants used to enter their answers. The use of these keys minimized large-scale hand movements. As before, participants were given 850 ms to make their initial decision, but in this experiment they entered their response by using the numeric keypad. The time that participants were given to answer a problem was a function of the strategy selected. When participants chose to retrieve, they were allotted 2.5 s to type in their full response. Pilot work established that 2.5 s was sufficient for participants to retrieve and enter a response but was not long enough for any of the participants to calculate the answer. When participants selected calculate, they were given 30 s to compute the answer. After executing the appropriate strategy, the participant was verbally informed by the computer of how many points he or she received, if any. The feedback was an auditory recording of the experimenter's voice. Tone and volume of the feedback were set to reward correct, on-time retrievals and admonish late responses. The participant was also informed if either of the deadlines was not met. The problem and correct answer were then presented for study. The study period was self-timed with a 2-s minimum. The participant began the next trial by pressing the space bar.

The reward structure was as follows: (a) Participants received 25 points when they selected the retrieval strategy, both strategy choice and strategy execution were on time, and the answer was correct. (b) Participants received 5 points when they selected to calculate, answered correctly, and met both deadlines. (c) If one of the two times to select the strategy or to give the answer was late, but the answer was correct, participants received 1 point. (d) Participants received 0 points when they met neither deadline or failed to answer correctly.

The participants taking part for course credit received \$0.001 per point plus a base fee of \$2.00. Participants who participated for money only received \$0.002 per point plus a base fee of \$5.00. As an additional performance incentive, all participants received an extra dollar if their point total for the 2 days exceeded the "current high score." The average paid for this experiment was \$5.41 for the credit participants and \$12.31 for the money-only participants.

After participants completed the first session, which was approximately 90 min in length, they were instructed to come back tomorrow "to do a slightly different task." When participants

returned for the second session, they were told that they would be doing the same task, but with new problems. They were quickly reminded of the task procedure and were then given the new set of problems. At the end of the second session, participants were asked whether they had noticed anything about today's problems and whether they had used any special strategies in deciding whether to choose retrieve or calculate. The second session was approximately 1 hr in length.

Design. The design of the problem set was similar to the one used in Experiment 1, with several modifications. On the first day, there were a total of 300 training trials. These were broken down into six blocks of 50 trials. The order of these trials was completely random. There were 16 problems, varying in frequency of presentation (high, medium, or low). A template for creating problems similar to those in Experiment 1 was used, except that there were no infrequently answered problems (i.e., all problems were always answered). The high-, medium-, and low-frequency problems were presented 27, 18, and 12 times, respectively. A high overall frequency of presentation for Day 1 problems was chosen to maximize transfer.⁹ Because two sets of 16 problems had to be created in this experiment, a slightly larger set of numbers was used from which to select operands: 12, 13, 14, 16, 17, 18, 19, 21, 23, 24, 26, 27, 28, 29, 31, 32, 34, 36, 37, 38.

On the second day, there was a total of 204 trials. Sixteen new training problems were created, again varying in frequency of presentation by using a template similar to that used for Day 1. The 16 new problems were selected from the same set of possible operands as in Day 1, with the constraint that the Day 2 operand pairings were completely novel. That is, if $a \# b$ was a problem on Day 1, then $a * b$, $b \# a$, and $b * a$ could not be used on Day 2. Other than this constraint, the newly selected pairs were chosen randomly. As a consequence, operands that were high frequency on Day 1 could be not present, high frequency, or low frequency on Day 2.

For Day 2, the high-, medium-, and low-frequency problems were presented 20, 10, and 5 times, respectively, resulting in 180 training trials. There were also 24 test (operator-switch) problems, 16 of which were from the Day 1 sample and 8 were from the Day 2 problems. All of the operator-switch problems were presented only once, ensuring that the participants would not know the answer to any of these problems. The 16 Day 1 test problems represented all of the Day 1 problems: 4 high frequency, 8 medium frequency, and 4 low frequency. The Day 2 operator-switch test problems were included as a replication of previous studies of short-delay familiarity (e.g., Reder & Ritter, 1992) by using the new interface. There were 8 such problems: 4 were operator-switches of high-frequency training problems, and 4 were operator switches of medium-frequency training problems. Low-frequency operator switches for Day 2 training problems were omitted to minimize the total number of test problems. Note that none of the previous day test problems involved exact repetitions of the previous day training problems; rather they were always presented with the operator switched. This design was deemed preferable to including exact repetition 24 delay test problems because the operator-switch problems were deemed a stronger test of the familiarity hypothesis, and the overall ratio of test problems to training problems on Day 2 had to be minimized.

The order of the training problems was completely random. The first 80 trials included only training trials. After the 80th trial, a test problem was presented once in every 5 trials. The location within each block was randomly selected with the constraint that the test problems could not occur consecutively. The relative order of the test problems was also completely random.

In summary, the goal of Experiment 2 was to establish that the familiarity-based feeling-of-knowing effects also occur in long-

term phenomena. On Day 1, participants were trained on one set of problems. On Day 2, participants began training on an entirely new set of problems. Unbeknownst to the participants, two kinds of test trials were inserted into the training trials. First, there were operator-switch problems based on problems from Day 2 (i.e., the same day). Second, there were operator-switch problems based on problems from Day 1 (i.e., the previous day). If there are long-term effects of problem familiarity, then participants should be more likely to select to retrieve for high-frequency previous day operator-switch problems than for medium- and low-frequency previous day operator-switch problems.

Results and Discussion

Four participants were dropped from the analyses: 3 participants did not finish the task because of a time constraint, and 1 participant did not complete the second day problems because of technical difficulties. Trials with very long strategy selection times (i.e., took longer than 2 s) were discarded (0.8% on Day 1 and 0.2% on Day 2). The analyses consisted of two parts: several global analyses of performance in this task, assessing the impact of the new interface, and more direct analyses of the hypotheses under study.

Operator differences. Tables 4 and 5 present summary statistics for performance on the training trials for both operators (* and #) on each day. As with Experiment 1, there was a small bias to select retrieve for sharp problems. The other difference between the operators was that the calculation times were longer for multiplication than for sharp, reflecting the relative difficulty of the algorithms—this difference may have been hidden in Experiment 1 because modulo arithmetic was used for that experiment and hence did not require complete calculation of the answers. However, because both operators exhibited similar behavior in all other respects, the data were collapsed over operators for all of the analyses presented below.

Strategy selection time and appropriateness of strategy selections. Participants were generally able to select a strategy before the 850-ms deadline. The mean strategy selection time was lower than 650 ms for both operators on Day 1 and lower than 550 ms on Day 2 (see Tables 4 and 5). Fewer than 16% of the strategy selections were late on Day 1 (greater than 850 ms), and this number dropped to fewer than 5% on Day 2.

In addition to generally making on-time selections, the appropriateness or accuracy of participants' selections was high. For multiplication problems, participants had a mean γ of .85 and d' of 1.62 on Day 1 and a mean γ of .81 and d' of 1.48 on Day 2. For sharp problems, participants had a γ of .68 and d' of 1.39 on Day 1 and a γ of .75 and d' of 1.34 on Day 2. Thus, with the new interface, participants were still able to make generally very accurate, on-time decisions. As with Experiment 1, we could not conduct d' analyses

⁹ Of course, these presentation rates are likely to be much lower than those typically experienced for arithmetic problems during primary education.

separately for the different problem types: Operator-switch problems occurred relative infrequently, and participants almost never knew the answer to those problems (participant means of 2.2% for previous-day operator switches and 3.1% for same-day operator switches vs. 35.6% for Day 2 training problems).¹⁰ However, as is shown in the next set of analyses, participants were just as likely to select retrieve for the operator-switch problems (to which they did not know the answer) as they were to select retrieve for the training problems (to which they frequently did know the answer).

The effect of practice on feeling of knowing. First, we wanted to ensure that the previously found familiarity effects would replicate, so analyses were conducted on the aspect of Experiment 2 that was most similar to previous studies: the Day 2 training and same-day operator-switch data. A logistic regression was conducted on the strategy selections for these trials (see Table 6). The frequency with which the bottom operand appeared, and the frequency with which both operands appeared, proved to be highly predictive ($|Z| = 2.99$, $p < .01$, and $|Z| = 6.05$, $p < .001$, respectively). However, whether the problem was a training problem or an operator-switch problem had no impact

Table 4
Participant Means and Standard Errors for All Problems on Day 1 of Experiment 2

Measure	Multiplication		Sharp	
	Calculation	Retrieval	Calculation	Retrieval
Strategy selected (%)				
<i>M</i>	59.2	40.8	58.8	41.2
<i>SE</i>	5.3	5.3	5.4	5.4
Strategy selection time (ms)				
<i>M</i>	636	687	638	611
<i>SE</i>	31	72	33	39
Late to choose strategy (%)				
<i>M</i>	14.3	19.0	16.7	12.3
<i>SE</i>	2.4	3.8	2.7	2.6
Correct answer times (ms)				
<i>M</i>	9,132	1,653	6,408	2,164
<i>SE</i>	546	217	430	425
Correct answer (%)				
<i>M</i>	87.8	80.5	88.4	78.5
<i>SE</i>	1.5	2.8	1.6	3.5
Incorrect choice of retrieval (% false alarms)				
<i>M</i>		22.1	26.7	
<i>SE</i>		3.6	4.3	
Both strategy choices				
$\gamma_{\text{feeling of knowing and knowing}}$				
<i>M</i>		.85	.68	
<i>SE</i>		.03	.11	
<i>d'</i>				
<i>M</i>		1.62	1.39	
<i>SE</i>		0.11	0.20	

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

Table 5
Participant Means and Standard Errors for Training Problems on Day 2 of Experiment 2

Measure	Multiplication		Sharp	
	Calculation	Retrieval	Calculation	Retrieval
Strategy selected (%)				
<i>M</i>	68.7	31.3	65.2	34.8
<i>SE</i>	4.8	4.8	5.1	5.1
Strategy selection time (ms)				
<i>M</i>	489	536	474	518
<i>SE</i>	30	25	27	25
Late to choose strategy (%)				
<i>M</i>	4.6	5.5	4.9	4.7
<i>SE</i>	1.6	1.5	1.5	1.7
Correct answer times (ms)				
<i>M</i>	8,604	1,972	5,713	1,954
<i>SE</i>	604	473	440	411
Correct answer (%)				
<i>M</i>	89.5	78.6	89.3	81.1
<i>SE</i>	1.8	4.0	2.0	3.5
Incorrect choice of retrieval (% false alarms)				
<i>M</i>		15.7	20.3	
<i>SE</i>		2.9	4.2	
Both strategy choices				
$\gamma_{\text{feeling of knowing and knowing}}$				
<i>M</i>		.81	.75	
<i>SE</i>		.04	.06	
<i>d'</i>				
<i>M</i>		1.48	1.34	
<i>SE</i>		0.15	0.15	

Note. Trials with late strategy selection were included only in the means of the strategy selection time. $N = 23$ for retrieval entries because 2 participants never selected to retrieve.

($|Z| = 0.27$, $p > .5$), suggesting that participants were completely fooled by these test problems. The results are qualitatively identical to those found by Reder and Ritter (1992)—participants made retrieve-compute decisions by using a partial-matching strategy. We now turn to the central issue explored in this experiment: the 24-hr delay results.

Long-term feeling of knowing. To test whether participants still thought problems were familiar after a 24-hr delay, we conducted a repeated-measures ANOVA on each participant's proportion of retrieval strategy selections for the high-, medium-, and low-frequency operator-switch test problems from the previous day. Figure 7a presents the mean rates for selecting retrieval, as well as data from the Day 2 training and the same-day operator-switch test problems for comparison. To make the time periods comparable, training data were only taken from the interval during which test trials were presented (i.e., Trials 81–204). The dotted line

¹⁰ Indeed, it is likely that participants did not know the answer to any operator-switch problems (never having seen them before) and correctly guessed the answer on a few problems.

Table 6
Results of Logistic Regression Predicting Strategy Selection for All Choice-on-Time Training and Operator-Switch Problems on Day 2 of Experiment 2

Variable	Range	Regression coefficient	SE	Z ^a
Intercept constant		-1.92	0.183	10.5*
Participant's own coded variable	0.02-4.13			0.09-5.62
Frequency of top operand	0-30	0.0035	0.0088	0.40
Frequency of bottom operand	0-30	0.026	0.0088	2.99*
Frequency with which both operands appeared together	0-20	0.12	0.019	6.05*
Type of problem	0-1	-0.05	0.182	0.27

^aComputed as (coefficient/SE_{coefficient}) in the regression.

* $p < .01$.

indicates the mean retrieval strategy selection rate for low-frequency training problems. Because participants are unlikely to be familiar with these low-frequency problems, this rate represents an estimate for the base rate with which retrieval is selected by mistake or from a bias to choose retrieve.

In agreement with the statistical analyses of the previous section, Figure 7a shows that there was no effect of switching operators on the same day problems with respect to participants' strategy selections. The effects of frequency of presentation on the previous day operator-switch test problems were in the expected direction, with participants being more likely to select retrieve for high-frequency than for medium-frequency test problems and being more likely to select retrieve for medium-frequency than for low-frequency test problems. However, the effects were small and not statistically significant, $F(2, 48) < 1$.

One factor that reduced the size of the effect was that several participants were at floor or ceiling on strategy selections (i.e., either rarely or always selecting to retrieve). To partial out the influences of this factor, we removed participants at floor and ceiling from the analyses. Participants were said to be at floor if they selected retrieve fewer than 10% of the time for the high-frequency training problems ($n = 3$). Participants were said to be at ceiling if they selected retrieve more than 50% of the time for low-frequency training problems ($n = 4$). Note that the data used to select which participants to remove were different from the data used in the statistical analysis of interest (i.e., independent criteria were used). Furthermore, these deleted participants do not contradict the familiarity hypothesis in anyway—their performance can easily be captured by using very high and very low response thresholds. Figure 7b presents the results with these participants removed. Once again, the effects were in the expected direction. Even the (relatively) low-frequency test problems demonstrated some resiliency to the 24-hr delay. The overall effect of frequency within the previous day test problems was marginally significant, $F(2, 32) = 3.02, p < .06$. Furthermore, the rate of selecting retrieve for the high-frequency previous-day test problems was significantly higher than the base rate of selecting retrieve, $F(1, 16) = 5.44, p < .05$.

The analyses presented thus far provided only marginal evidence for long-term effects. To further investigate these

effects, we redid the analyses as linear regressions. However, rather than using presentation frequency as a predictor, we selected a different measure of item familiarity: the rate at which the retrieval strategy was selected for training problems on the previous day. Because participants often reported using idiosyncratic patterns to make retrieve-compute decisions, this measure is likely to be a more sensitive measure of item familiarity as it included both problem frequency and problem-idiosyncratic features. Previous day retrieval selection rate was a significant predictor of (operator-switch) test retrieval selection rates ($r = .707, p < .0001$). Thus, there is strong evidence for long-term effects.

However, it may be that these long-term effects were due to participants getting an early read on the answer rather than simple familiarity with the problem. To test among these alternative accounts, we conducted a multiple regression using the rate of retrieval strategy selections and the rate of actual retrievals (i.e., correct answers given in less than 2 s). If the early-read hypothesis is correct, then the rate of actual retrievals should be a strong predictor of test problem choices. By contrast, if the familiarity hypothesis is correct, then previous-day strategy selections should be the only independent predictor. The results of the multiple regression supported the familiarity hypothesis: Previous-day strategy selections predicted test problem strategy selections ($\beta = 0.75$), $F(1, 397) = 62.40, p < .0001$, whereas the previous day rate of actual retrievals did not ($\beta = -0.02$), $F(1, 397) < 1, p > .85$.

When we turn to the modeling of these data, we argue that the attenuation of the frequency of presentation effects found for the previous day test problems was due to the decay of long-term strength over the delay, just as the SAC model predicts. However, an alternative explanation for this phenomenon is that participants might have become aware that all problems from the previous day were operator-switch problems and that they should select to calculate whenever they recognize a problem from the previous day. At an intuitive level, this alternative explanation seems implausible because participants could not know that trials from the previous day would all be operator switches until the task was almost over, assuming that they even recognized that the trials were operator switches of previous-day problems. However, to provide more concrete evidence against this

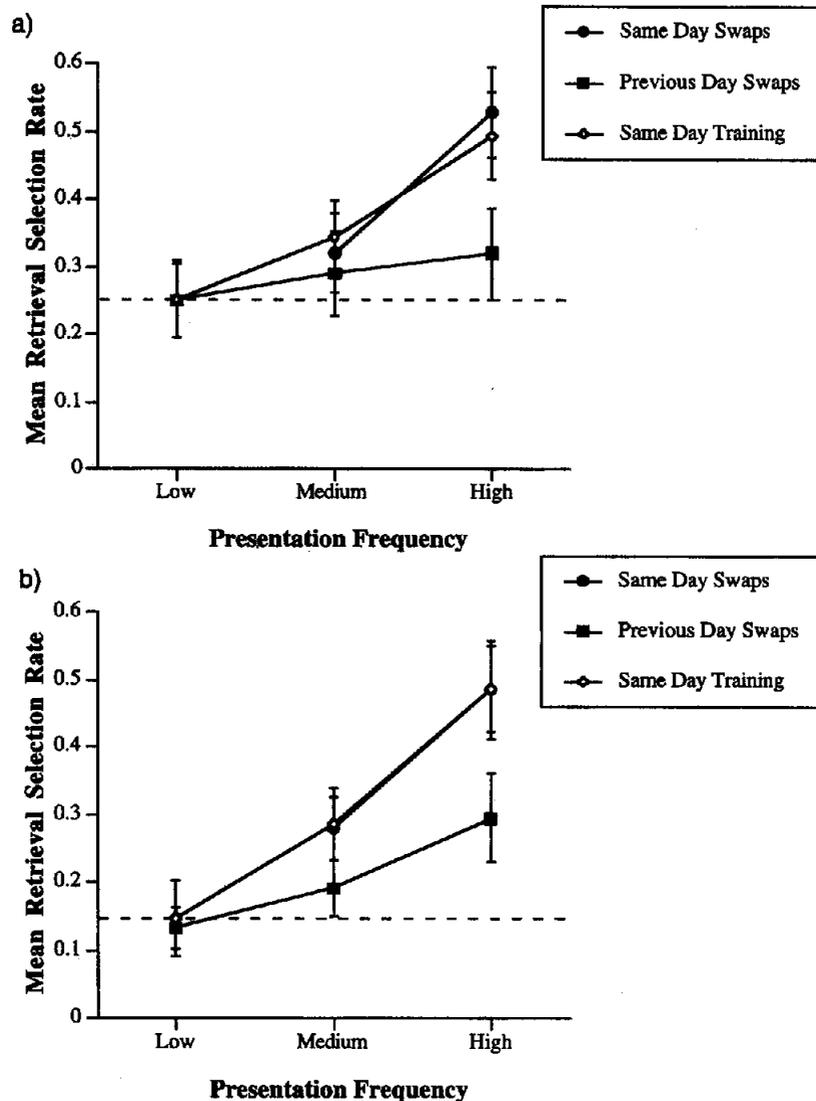


Figure 7. For (a) all participants and (b) with participants at floor and ceiling removed, the mean rate of selecting retrieval (and standard error) in Experiment 2 at each frequency level for same and previous day operator-switch test problems and for comparable-time training trials.

alternative account, we analyzed the awareness data, in which we asked participants at the end of the second day whether they had noticed anything special about the second-day trials. Ten of the 25 participants reported being aware that there were operator-switch trials intermixed with the regular trials. We then analyzed whether these aware participants were less likely to select retrieve overall, or whether they were less likely to select retrieve for the previous-day operator-switch trials. There were no significant effects of awareness on either factor, $F(1, 23) < 1, ps > .6$ for both. In the case of overall retrieve rates, if anything, aware participants were slightly more likely to select retrieve overall ($M = 35.8\%$, $SE = 10.0$) than were nonaware participants ($M = 31.7\%$, $SE = 6.2$). Similarly, if anything, aware participants were slightly more likely to select to retrieve for the

previous-day operator-switch trials ($M = 30.0\%$, $SE = 5.6$) than were the unaware participants ($M = 27.8\%$, $SE = 4.6$). Thus, it is unlikely that the attenuation of the presentation frequency effects after a 24-hr delay was due to the participants using a metastrategy of not selecting retrieval for trials recognized to have occurred on the previous day. Rather, it is more likely that the attenuation was due to decay of long-term strength.

In summary, the results of Experiment 2 demonstrate that there are long-term feeling-of-knowing effects on strategy selection and that rapid feeling of knowing is not simply a short-term phenomenon. Furthermore, problem familiarity rather than early reads on the answer underlies these strategy selections, even in the long-term case. To see whether the exact size of the long-term effects are predicted by the

model, we now present the simulations of Experiment 2 results.

Experiment 2 Simulations

As with the fit to the other data sets, we compared the SAC model's predictions of each participant's actual retrieve-compute decisions on the basis of the actual frequency of problems seen by that participant. Again, all of the same parameter values from the previous simulations were used for all but the two participant-specific parameters (i.e., the participant's threshold, and whether they used the never-retrieve rule for an operator). Further, the two participant-specific parameters were held constant across the simulations of Experiment 2 data (e.g., the same values were used for Day 1 and Day 2 simulations). The best fitting participant thresholds ranged from 35 to 180, with a mean threshold of 97.2 ($SD = 35.6$).

As in the simulations of the previous experiments, for simplicity, trials were used rather than time to represent the unit of learning and forgetting. Therefore, the size of the delay had to be estimated in number of trials. On Day 1, participants completed approximately 300 trials in 90 min. Therefore, 24 hr should be the equivalent of 4,800 trials. Consequently, the 24-hr delay was simulated by decaying all node strengths and link strengths by 4,800 trials.

As with the other data sets, SAC fit the training data quite well. The model accounted for 76% of the variance in the individual participant's strategy selections for the Day 1 data and 63% of Day 2 data (see Figures 8a and 8b). Furthermore, the slopes of the best fitting lines were not significantly different from one in either case (slope = 1.02), $t(287) = 0.50$, $p > .6$ for Day 1, and slope = 1.07, $t(172) = 1.09$,

$p > .25$ for Day 2. However, the intercepts deviated slightly from zero—statistically significantly in the case of Day 1 (intercept = -0.051), $t(287) = 2.79$, $p < .01$, and very marginally in the case of Day 2 (intercept = 0.038), $t(172) = 1.35$, $p > .15$. It seems that the SAC model slightly overpredicted retrieval rates on Day 1 and underpredicted retrieval rates on Day 2. These effects may reflect shifting thresholds across the two experiment sessions.

These preceding model fits did not depend significantly on the use of the never-retrieve rule: Only 1 participant was classified as using this rule, and the overall correlation between predicted and actual strategy selection rates remained unchanged when the never-retrieve rule was not used in the simulations.

One important result of Experiment 2 was a replication of the Reder and Ritter (1992) findings on the effect of operator switches on strategy selection. That is, as the SAC model predicts, participants were just as likely to select retrieve for the operator-switch test problems as for the original training problems. Because each participant received very few same-day operator-switch problems, only a small number of probability ranges could be used, and there were few observations contributing to each participant's proportions within each range. Given the large amount of noise in the data due to low n s, the fit of the model was adequate, accounting for 46% of the variance in the individual participant's proportion of retrieval strategy selections (see Figure 9a). The slope of the best fitting line was not significantly different from one (slope = 0.823), $t(32) = 1.07$, $p > .25$, and the intercept was only slightly above zero (intercept = 0.139), $t(32) = 2.17$, $p < .05$.

The major result of Experiment 2 was the discovery of

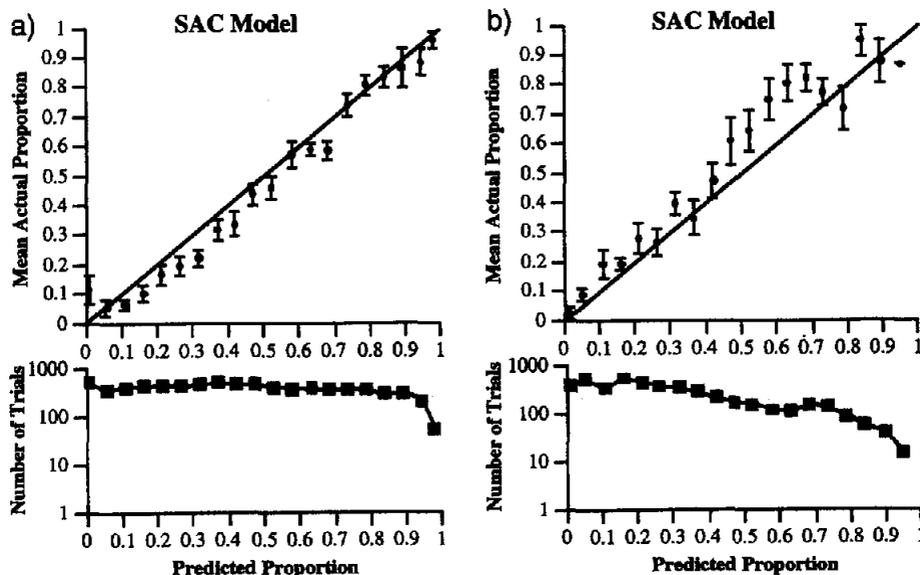


Figure 8. For (a) Day 1 training problems and (b) Day 2 training problems, in Experiment 2, mean actual proportion of retrieval strategy selections (and standard error) as a function of group predicted proportions of retrieval strategy selections, along with the number of trials in each grouping. SAC = source of activation confusion.

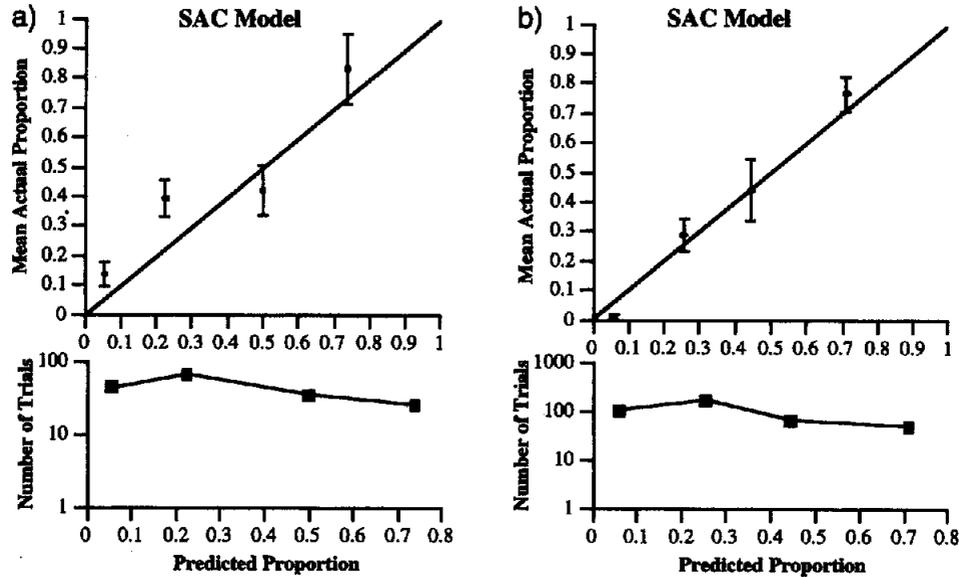


Figure 9. For (a) same day operator-switch test problems and (b) previous day operator-switch test problems, in Experiment 2, mean actual proportion of retrieval strategy selections (and standard error) as a function of group predicted proportions of retrieval strategy selections, along with the number of trials in each grouping. SAC = source of activation confusion.

long-term familiarity effects on strategy selections. The SAC model predicts that there should be long-term effects. However, it is not necessarily true that the SAC model will be able to account for the absolute magnitude of these long-term effects because their magnitude depends strongly on the details of the underlying memory model.

One method for assessing the SAC model's long-term predictions is to compare the predicted and actual strategy selection rates in the same fashion that the other fits were evaluated: grouped by model prediction rates. With this method of evaluation, the SAC model performed quite well, accounting for 71% of the variance in the individual participants' selections (see Figure 9b). Furthermore, the slope of the best fitting line did not deviate significantly from one (slope = 1.19), $t(32) = 1.38$, $p > .15$, nor did the intercept deviate significantly from zero (intercept = -0.047), $t(32) = -0.96$, $p > .3$.

Another method for assessing the SAC model's long-term predictions is to compare the predicted and actual effects of frequency of presentation directly. That is, rather than plotting the observed data as a function of the model predictions, one can plot the observed data as a function of conditions and also plot the predicted values as a function of these conditions. This more typically used method provides a direct comparison of the simulation results to the empirical findings. Figure 10 presents the actual rate of retrieval strategy selections for the previous-day operator-switch test problems (first presented in Figure 7a) with the predicted retrieval strategy selection rates for each frequency level. The actual and predicted strategy selection rates were generally quite close. In particular, the predicted mean was always within one standard error of the actual mean. Thus,

any deviations between predicted and actual could have been due to noise in the data.

General Discussion

The two experiments presented in this article have generated further support for the view that rapid feeling of knowing and strategy selection are based on features of the question or problem statement. In particular, the experiments provided evidence against two alternative hypotheses:

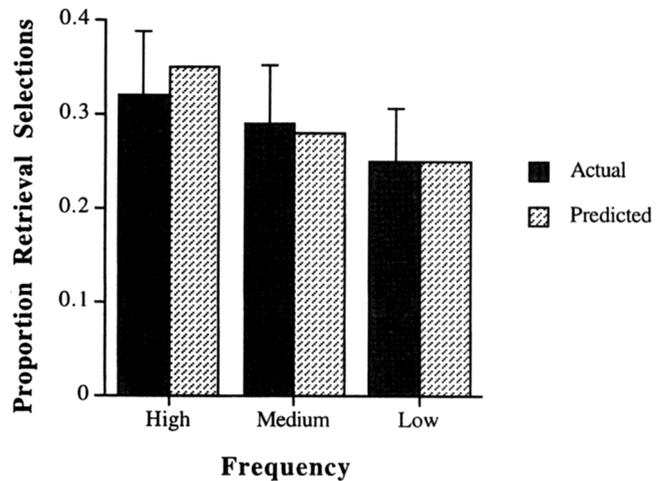


Figure 10. Mean actual (with standard error) and predicted rates of retrieval strategy selections for each level of presentation frequency for previous day operator-switch test pairs in Experiment 2.

(a) the hypothesis that participants were basing their decisions on an early-read of some (possibly incorrect) answer, and (b) the hypothesis that the previously found problem-familiarity effects would not generalize to situations with long delays.

The SAC model of feeling of knowing and strategy selection has been applied to three data sets. It is the first computational model of feeling of knowing and strategy selection to be fit to feeling-of-knowing data in a rigorous fashion. For each experiment that we simulated, our fit to data was very good—both the relative ordering of retrieval strategy selections on various problems and the absolute magnitude of selection rates were fit well by the SAC model. These strong fits are especially impressive given that the same basic parameter values were used in all of the simulations.

Although we accounted for large percentages of the variance in the data, we did not account for all of it. This observation leads to the question: What was the source of the remaining variance? We believe that, in large part, the answer is noise. The model fits that we presented were at the individual participant level, rather than the more typically used across-participants aggregate level. Consequently, many of the observed values were based on relatively few observations. Because the dependent measure was a binary variable, the resulting proportions were highly unstable for low *ns*. When we fit the SAC model by aggregating across participants (see Reder & Schunn, 1996), we found that in all cases the model could account for over 95% of the variance in strategy selections. However, because the SAC model was yoked to each individual participant's stimuli, and the individual level is a more difficult one to predict, we felt that tests of the model at the individual participant level were the more stringent tests.

To provide some contrast to the SAC model, we tested two alternative models. We found that a model using only participant thresholds could not account for nearly as much of the variance as the SAC model. In other words, the good fits that we found could not be attributed to having a free parameter associated with each participant. Furthermore, such a simple model could not account for the effects of any variable on strategy selection (e.g., the effects of problem presentation frequency). Similarly, a simple base-rate model, while able to account for a significant proportion of the variance, did not provide as good a fit overall, despite having many free parameters. Furthermore, this base-rate model could not account for important portions of the data (e.g., the operator-switch data). Given these basic limitations of the alternative models, we only presented these comparisons for the fits to the Reder and Ritter (1992) data. However, as one would expect given these important limitations, testing the alternative models by using data from Experiments 1 and 2 produced the same findings: The SAC model accounted for significantly more variance.

Another important feature of the SAC model is that there are no ad hoc assumptions, and that this same model is being used by Reder to account for other phenomena such as cognitive illusions and word frequency effects (e.g., Kamas & Reder, 1994; Reder, Nhoyvanyisvong, Schunn, Angstadt

& Hiraki, 1996; Reder & Gordon, 1996). Although the never-retrieve rule might be considered an ad hoc assumption, the SAC model was still able to produce good fits to data without this feature. The type of activation-based model that we used is part of a powerful class of such models that have been used to account for a wide range of memory phenomena, including the shape of learning functions (e.g., Anderson, 1993; Anderson & Schooler, 1991; McClelland & Rummelhart, 1986), the shape of forgetting functions (e.g., Anderson & Schooler, 1991), and fan effect phenomena (e.g., Anderson, 1983; Reder & Ross, 1983).

Another strength of the SAC model, as compared with competing models of the retrieve-compute selection process, is that in addition to doing a much better job of accounting for the effects found in our experiments, the SAC model's assumptions are consistent with and can capture many other basic memory phenomena. Other recent models of the retrieve-compute selection process such as Siegler and Shipley's (1995) ASCM or Logan's (1988) Instance Theory do not seem equipped to account for basic cognitive phenomena such as forgetting.

In this article we have focused on one function of feeling of knowing—strategy selection. Other functions of feeling of knowing include regulating search length during retrieval (Gruneberg, Monks, & Sykes, 1977; Lachman & Lachman, 1980; Nelson, Gerler, & Narens, 1984; Reder, 1987, 1988; Ryan, Petty, & Wenzlaff, 1982) and adjusting memory trace strengths during learning (Metcalf, 1993, 1994). The SAC model could potentially be extended to account for these other functions of feeling of knowing as well by applying the SAC model's activation values to these other processes.

The SAC model of feeling of knowing and strategy selection that we have presented is based on a spreading-activation semantic network. Yet this was not the only way that we could have fit the data. An appropriately specified convolution or compound cue model could also be fit to these data. The goal of our modeling attempt was not to demonstrate the unique virtues of a spreading-activation model. Instead, we sought to demonstrate how a widely applied, general model of memory could be extended to account for strategy selection and feeling-of-knowing judgments. An important aspect of the SAC model—that feeling-of-knowing is driven by familiarity with components of the problem-question—is a relatively unique contribution. The only similar model of feeling of knowing is Metcalfe's composite holographic associative recall model (CHARM) (Metcalf, 1993, 1994; Metcalfe, Cottrell, & Mencl, 1993), which is based on a convolution vector model of memory. It is difficult to evaluate the relative contribution of the SAC model versus Metcalfe's CHARM model. Her model also posits that features of the problem influence feeling of knowing, but unlike SAC, it has never formally been applied or fit to an intricate set of data.

As indicated earlier, the SAC model bears similarity to other semantic network models of cognition; however, it also differs from these other models in important ways. In particular, many of these models (e.g., Anderson's, 1983, ACT) do not address the issue of strategy selection. Even the current ACT-R (Anderson, 1993), which does contain a

theory of strategy selection, fails to recognize that features of the problem affect strategy choice. The SAC model also differs significantly from other kinds of models of strategy selection, which have typically assumed that participants always attempt to retrieve first and only attempt other strategies on retrieval failures (e.g., LeFevre, Greenham, & Waheed, 1993; Siegler, 1987, 1988; Siegler & Shrager, 1984).

We take the findings of this article to support the view that people can and do decide between trying to retrieve an answer from memory and trying to compute the answer, and they make this decision before trying to retrieve the answer. That is, we argue that people do not always try to retrieve first. Furthermore, the decision process cannot be captured with a simple horse race between retrieval processes and computing processes. If the decision process were a horse race, then factors affecting the retrievability of the answer from memory and the speed of execution of the computing strategy should be the only factor affecting the decision process (see also Lemaire & Reder, 1996). However, Jameson, Narens, Goldfarb, and Nelson (1990) have shown that priming of the answer affects retrievability of the answer but does not affect feeling-of-knowing judgments, and Reder (1987) found that priming of question statements increased decisions to retrieve but did not affect retrievability of the answer. These findings are in direct conflict with the fundamental assumption of many models of cognition (e.g., Logan's, 1988, Instance Theory; Anderson's, 1983, ACT theory) that memory retrieval is an automatic process not subject to adaptive strategy selection processes.

Another feature of the SAC model is that it shows how feeling of knowing is used as a component of strategy selection—we have provided a solution to the computational conundrum associated with assessing question familiarity to make retrieve–compute decisions. We illustrated with our SAC model that problem familiarity can be used to guide strategy selection because the assessment inherently occurs before retrieval. Empirically we have shown that it is an accurate predictor of retrieval success despite the fact that it can be subverted; that is, it is a heuristic that is not based on an earlier read of the answer.

How well do our findings generalize to other kinds of problems and tasks? It might be argued that people do not make the strategic decision between computing and retrieving for very simple problems like retrieving someone's phone number or answering simple arithmetic problems like $2 * 2$. However, there are several reasons to believe that SAC applies to such situations as well. In the case of trying to retrieve someone's phone number, one must decide whether to search the phone book or search memory first. Even for simple arithmetic problems, recent research has found that a majority of adults (including university students) occasionally use a compute strategy for some single digit addition and multiplication problems (LeFevre, Sadesky, & Bisanz, 1996). That is, even very simple problems are not exempt from the strategic decisions. Furthermore, the fact that people usually retrieve answers for very simple answers is consistent with this model—SAC predicts that people will

select to retrieve for highly familiar problems, which such simple problems usually are.

Another potentially problematic case for our familiarity-driven strategy selection model is the use of the never-retrieve rule. Although we were easily able to model such behavior with the addition of this simple rule, the use of a rule of this type is a clear exception to familiarity-driven strategy selection. This contrast between the familiarity-driven heuristics and the never-retrieve rule is but one instance of a more general distinction between the two types of factors that influence strategy selection: features contained in the question or problem (e.g., the problem familiarity), and features contained outside the problem or question (e.g., the history of success of a particular strategy or operator). Reder (e.g., Miner & Reder, 1994; Reder, 1987, 1988; Reder & Schunn, 1996) has labeled this the distinction between intrinsic and extrinsic factors. Relating our current modeling enterprise to this distinction, feeling of knowing is an intrinsic factor that combines with extrinsic factors to produce strategy selections (cf. Reder, 1987). For example, variables that influence feeling of knowing are integrated with other factors such as “strategy X is working well” or “always select retrieve and try to ‘beat the clock’ when the operator is addition” (Reder & Ritter, 1992). Even outside of experimental settings, it is likely that people use other metarules when making a retrieve–compute decision. For example, research has found that people quickly compute the answer to $n * 0$ problems, with the answer time being independent of the size of n (Ashcraft, 1982). It is worth noting that these examples of other factors influencing strategy selection are consistent with the general approach of our SAC model that states that features of the problem but not of the answer influence strategy selection. Future computational work should be directed toward extending the SAC model in these directions.

We believe the primary contributions of the current article to be threefold. First, using data from Experiments 1 and 2 to rule out several very plausible alternative explanations, we have provided much stronger evidence for the claim that people typically make retrieve–compute strategy decisions on the basis of familiarity with the question or problem statement. Second, we have provided a mechanistic account of how such a decision process might work. Finally, we have demonstrated that it is possible to provide a detailed, quantitative account for individual participants' strategy selections and feeling-of-knowing decisions.

References

- Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, 6, 451–474.
- Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (1993). *Rules of mind*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, 2, 396–408.

- Ashcraft, M. H. (1982). The development of mental arithmetic: A chronometric approach. *Developmental Review*, 2, 213-236.
- 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.
- Connor, L. T., Balota, D. A., & Neely, J. H. (1992). On the relation between feeling of knowing and lexical decision: Persistent subthreshold activation or topic familiarity? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(3), 544-554.
- 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 31st annual meeting of the Psychonomics Society, New Orleans, LA.
- 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.
- 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.
- 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.
- Kamas, E., & Reder, L. M. (1994). The role of familiarity in cognitive processing. In E. O'Brien, & R. Lorch (Eds.), *Sources of coherence in text comprehension: A festschrift in honor of Jerome L. Myers* (pp. 177-202). Hillsdale, NJ: Erlbaum.
- Lachman, J. L., & Lachman, R. (1980). Age and 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. 285-311). Hillsdale, NJ: Erlbaum.
- Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence*, 33, 1-64.
- LeFevre, J., Greenham, S. L., & Waheed, N. (1993). The development of procedural and conceptual knowledge in computational estimation. *Cognition and Instruction*, 11, 95-132.
- LeFevre, J., Sadesky, G. S., & Bisanz, J. (1996). Selection of procedures in mental addition: Reassessing the problem-size effect in adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 216-230.
- Lemaire, P., & Reder, L. M. (1996). *What affects strategy selection in arithmetic? An example of parity and five effects on product verification*. Manuscript submitted for publication.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95(4), 492-527.
- Lovett, M. C., & Anderson, J. R. (in press). History of success and current context in problem solving: Combined influences on operator selection. *Cognitive Psychology*.
- McClelland, J. L., & Rummelhart, D. E. (1986). *Parallel distributed processing: Explorations in the microstructures of cognition, Vol. 2: Psychological and biological models*. Cambridge, MA: MIT Press.
- Metcalf, J. (1993). Novelty monitoring, metacognition and control in a composite holographic associative recall model: Implications for Korsakoff amnesia. *Psychological Review*, 100, 3-22.
- Metcalf, J. (1994). A computational modeling approach to novelty monitoring, metacognition, and frontal lobe dysfunction. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 137-156). Cambridge, MA: Bradford.
- Metcalf, J., Cottrell, G. W., & Mencl, W. E. (1993). Cognitive binding: A computational-modeling analysis of a distinction between implicit and explicit memory. *Journal of Cognitive Neuroscience*, 4, 289-298.
- Miner, A., & Reder, L. M. (1994). A new look at feeling-of-knowing: Its metacognitive role in regulating question answering. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 47-70). Cambridge, MA: Bradford.
- Nelson, T. O. (1984). A comparison of current measure 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., & 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. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, 95(3), 385-408.
- Reder, L. M. (1979). The role of elaborations in memory for prose. *Cognitive Psychology*, 11, 221-234.
- 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.
- Reder, L. M., & Gordon, J. S. (1996). Subliminal perception: Nothing special cognitively speaking. In J. Cohen & J. Schooler (Eds.), *Cognitive and neuropsychological approaches to the study of consciousness* (pp. 125-134). Hillsdale, NJ: Erlbaum.
- Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Angstadt, P., & Hiraki, K. (1996, November). *Modeling word frequency effects in a continuous remember/know judgment paradigm*. Paper presented at the meeting of the Psychonomics Society, Chicago, IL.
- Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with questions terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 435-451.
- Reder, L. M., & Ross, B. H. (1983). Integrated knowledge in different tasks: The role of retrieval strategy on fan effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9, 55-72.
- Reder, L. M., & Schunn, C. D. (1996). Metacognition does not imply awareness: Strategy choice is governed by implicit learning and memory. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 45-77). Hillsdale, NJ: Erlbaum.

- Ryan, M. P., Petty, C. R., & Winzloff, R. M. (1982). Motivated remembering efforts during tip-of-the-tongue states. *Acta Psychologica, 51*, 137-157.
- Schreiber, T. A., & Nelson, D. A. (1995). *Feelings of knowing are sensitive to the activation of neighbouring concepts*. Manuscript submitted for publication.
- Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1074-1083.
- Siegler, R. S. (1987). Strategy choices in subtraction. In J. Sloboda & D. Rogers (Eds.), *Cognitive processes in mathematics* (pp. 81-106). Oxford, England: Oxford Books.
- Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General, 117*, 258-275.
- Siegler, R. S., & Jenkins, E. (1988). *How children discover new strategies*. Hillsdale, NJ: Erlbaum.
- Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In G. Halford & T. Simon (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 31-76). New York: Academic Press.
- Siegler, R. S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do. In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229-293). 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, L., & Meyer, D. E. (1987). Activation and metacognition of inaccessible stored information: Potential bases for incubation effects in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 187-205.

Received January 25, 1995

Revision received May 13, 1996

Accepted May 14, 1996 ■

Low Publication Prices for APA Members and Affiliates

Keeping you up-to-date. All APA Fellows, Members, Associates, and Student Affiliates receive—as part of their annual dues—subscriptions to the *American Psychologist* and *APA Monitor*. High School Teacher and International Affiliates receive subscriptions to the *APA Monitor*, and they may subscribe to the *American Psychologist* at a significantly reduced rate. In addition, all Members and Student Affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

Essential resources. APA members and affiliates receive special rates for purchases of APA books, including the *Publication Manual of the American Psychological Association*, and on dozens of new topical books each year.

Other benefits of membership. Membership in APA also provides eligibility for competitive insurance plans, continuing education programs, reduced APA convention fees, and specialty divisions.

More information. Write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242.