

Production, verification, and priming of multiplication facts

JAMIE I. D. CAMPBELL
Carnegie-Mellon University, Pittsburgh, PA

In the arithmetic-verification procedure, subjects are presented with a simple equation (e.g., $4 \times 8 = 24$) and must decide quickly whether it is true or false. The prevailing model of arithmetic verification holds that the presented answer (e.g., 24) has no direct effect on the speed and accuracy of retrieving an answer to the problem. It follows that models of the retrieval stage based on verification are also valid models of retrieval in the production task, in which subjects simply retrieve and state the answer to a given problem. Results of two experiments using single-digit multiplication problems challenge these assumptions. It is argued that the presented answer in verification functions as a priming stimulus and that on "true" verification trials the effects of priming are sufficient to distort estimates of problem difficulty and to mask important evidence about the nature of the retrieval process. It is also argued that the priming of false answers that have associative links to a presented problem induces interference that disrupts both speed and accuracy of retrieval. The results raise questions about the interpretation of verification data and offer support for a network-interference theory of the mental processes underlying simple multiplication.

Much of the modern experimental research on memory for simple arithmetic combinations has employed one or the other of two tasks. In the *verification* task, subjects are presented with an arithmetic equation (e.g., $3 + 8 = 11$, $4 \times 8 = 24$) and must indicate as quickly as possible whether the equation is true or false. This can be contrasted with the *production* task, in which subjects simply generate and state the answer to a given problem. The verification task has methodological advantages over the production task. First, the recording of responses can be completely automated for verification. With a true/false-button box connected to a computer that presents the problems and times the responses, many trials can be collected in a short period of time. The experimenter need not be present during testing. Moreover, several studies have shown that verification times can be manipulated as a function of properties of false answers. From these effects, models of the mental processes underlying arithmetic performance can be inferred (e.g., Krueger & Hallford, 1984; Parkman & Groen, 1971; Stazyk, Ashcraft, & Hamann, 1982; Zbrodoff & Logan, 1986). For the production task, accurate estimates of response times (RTs) require the use of a voice-activated relay that detects the onset of the subject's verbal response and stops a timer. Inevitably, some trials are lost because

the response fails to trigger the relay or because extraneous sounds stop the timer prematurely. An experimenter must be present to record each response or at least to record whether each response was correct or incorrect. Consequently, fewer production trials than verification trials can be collected in the same period of time. Furthermore, the production task does not afford the false-answer manipulations that verification does.

Despite the conveniences of the verification task, there remains a basic question: Do the verification and production tasks provide equivalent characterizations of arithmetic skills and processes? If one uses verification to study arithmetic as it is normally performed, the verification task must reflect accurately the retrieval or computational processes that operate in arithmetic production. Furthermore, the arithmetic-verification procedure is emerging as a potentially important paradigm for investigating general theoretical issues of cognition (e.g., the relation between the surface structure of a symbol and the processing of its content [Gonzalez & Kollers, 1982]; automaticity [Zbrodoff & Logan, 1986]). The interpretation of verification data, however, often depends upon assumptions about the arithmetic processes it incorporates and assumptions about how the verification component interacts with these processes. Consequently, it is important to determine whether theories of cognitive arithmetic based on production data can be assumed to apply equally to the verification task.

Production Versus "True" Verification

One view of arithmetic verification is that it normally comprises four discrete stages (e.g., Ashcraft, 1982, 1986; Parkman & Groen, 1971). The first stage is the initial encoding of the problem. The second stage consti-

This research was supported by a postdoctoral fellowship from the Natural Sciences and Engineering Research Council of Canada. I thank Mark Ashcraft, Derek Besner, Neil Charness, Alice Healy, Danica Lavoie, Pierre Jolicoeur, Bob Siegler, and an anonymous reviewer for helpful comments on an earlier version of the paper. A portion of these data was presented at the meeting of the Canadian Psychological Association, Toronto, Canada, June 1986. The author is now at the Department of Psychology, University of Western Ontario, London, Ontario N6A 5C2, Canada.

tutes the actual arithmetic component and may involve retrieval or computation of a response to the presented problem. The third stage involves comparison of the retrieved answer with the presented answer and a decision about whether or not they match. The fourth stage is the execution of the "true" or "false" response. In the present paper, the central question is whether the performance characteristics of the arithmetic stage are the same in both the verification task and the production task. The two tasks possibly might be very different in this respect. In an analysis of the sentence-verification task, a procedure analogous to arithmetic verification, Reder (1982) proposed that making plausibility judgments about sentences can be more efficient than a direct-retrieval strategy. Similarly, in arithmetic verification, instead of retrieving an answer and comparing it with the one presented, subjects might first judge the plausibility of an equation by checking whether the presented answer is approximately correct in terms of magnitude (cf. Ashcraft & Stazyk, 1981) or check whether the problem's operands are consistent with the odd-even status of the presented answer (e.g., in multiplication, if either multiplier is even, then the correct answer must also be even) (see Krueger, 1986, and Krueger & Hallford, 1984, for evidence that subjects can use such information efficiently). Thus, it is possible that on some trials there is no answer-generation stage in arithmetic verification.

Reder (1982) argued, however, that the choice of strategy in verification tasks depends upon accessibility, so that when a fact has been encoded recently or is represented by a very strong memory trace, a direct-retrieval strategy is promoted. Most educated adults can generate correct answers to any of the single-digit addition or multiplication problems in 1 sec or less (e.g., Miller, Perlmutter, & Keating, 1984). The ready accessibility of correct answers probably encourages a retrieve-compare strategy in arithmetic verification (Krueger, 1986). Indeed, there is evidence that the processing of simple arithmetic problems may be partially automatic in adults (LeFevre, Bisanz, & Mrkonjic, 1986; Zbrodoff & Logan, 1986), which suggests that retrieval operations are initiated regardless of subjects' intentions. Although it is true that subjects could base performance on a judgment of global familiarity or on plausibility judgments, except when experimental conditions encourage subjects to do otherwise, it is probably the case that a retrieve-compare strategy is dominant in adults' arithmetic verification.

Even allowing that both the production and verification tasks involve essentially the same arithmetic components (usually assumed to be associative recall in older children and adults, but see Baroody, 1984, 1985), it is well known that performance on false verification trials is affected by particular qualities of the false answer presented. For example, if the difference between a false answer and the correct answer (the *split*) is large, subjects can reject an answer as false more rapidly than when the numerical magnitudes are similar (e.g., Ashcraft & Stazyk, 1981; Stazyk et al., 1982). Similarly, if the false

answer would be correct via another arithmetic operation (e.g., $3 \times 5 = 8$ or $3 + 5 = 15$), verification times are slowed (Hamann & Ashcraft, 1985; Winkelman & Schmidt, 1974; Zbrodoff & Logan, 1986). Thus, false verification includes effects that may mask the arithmetic component of the task. This means that if we want an unobstructed view of the arithmetic retrieval stage using the verification task, we must rely on trials in which the presented answer is correct. Given the above model, even on true verification trials, of course, a comparison-decision stage is still required. Ashcraft (1982, 1986) argued that for true trials, however, the decision stage simply adds a constant amount of time and therefore would not distort effects due to retrieval or computational operations. It follows that models of arithmetic retrieval based on true verification (e.g., Ashcraft, 1982, 1986; Ashcraft & Stazyk, 1981; Stazyk et al., 1982) are also valid models of retrieval performance in the production task.

Baroody (1984, 1985) argued, however, that extrapolation of verification performance to production may be more complicated than has been acknowledged and that even true-verification data may include effects that are unique artifacts of the verification procedure. To evaluate the comparability of the tasks, Ashcraft, Fierman, and Bartolotta (1984) tested production and verification performance in mental addition for first-graders, fifth-graders, and college students (Experiment 1) and for first-through third-graders (Experiment 2). Their main analyses focused on tests of whether the *problem-size* effect was the same in the two tasks. It is well established that, in general, problem difficulty in single-digit addition and multiplication increases with the numerical magnitude of the problem in both verification (e.g., Ashcraft & Stazyk, 1981; Stazyk et al., 1982) and production tasks (e.g., Miller et al., 1984). The effect usually is analyzed by correlating correct RT across problems with one or a number of different problem-size variables. Typically, these include (e.g., for 6×9) the minimum operand (6), the maximum operand (9), the sum (15) or squared sum (225) of the operands, or the product of the operands (54). Although a number of different factors have been proposed to explain the relationship between problem size and difficulty (e.g., Ashcraft, 1986; Campbell & Graham, 1985; Seigler, 1986a), it appears to be universally accepted that the effect arises in the arithmetic retrieval or computation stage as opposed to during encoding or response-execution operations. Thus, differences in speed and accuracy across problem size are taken to reflect directly the arithmetic component.

Ashcraft et al. (1984) reported that, for the fifth-graders and college students, the sum squared was the best predictor of true-verification RTs, but the minimum addend was best for production RTs. This suggests that the tasks may differ, at least for more experienced subjects (Baroody, 1984). As Miller et al. (1984) pointed out, however, the problem-size variables are all highly intercorrelated, and thus the significance of one as opposed to another emerging as the best predictor should be interpreted cautiously.

Furthermore, Ashcraft et al. found that when RTs were correlated with the ranked difficulty measure reported by Wheeler (1939), the correlations were higher than those provided by problem-size variables for both the production and the verification tasks, and that the slopes were virtually identical beyond the first grade. Analyses of variance (ANOVAs) with tasks, grades, and problem size (sum < 10 versus ≥ 10) as factors confirmed that, although true verification consistently required longer RTs than did production, a result attributable to the additional comparison-decision stage in verification, there were no significant task \times problem-size interactions beyond the first grade. Ashcraft et al. concluded that, for addition, the two tasks provide essentially the same profiles of the retrieval process.

It should be noted, however, that problem-size variables and the Wheeler (1939) difficulty index provide only approximate indexes of problem difficulty. Usually more than 35% of the variance in difficulty across problems is not accounted for by problem size (Campbell & Graham, 1985). Even the Wheeler difficulty measure used by Ashcraft et al. (1984) accounted for only 61% of the variance in RT across problems (averaging over tasks, grades, and experiments). Consequently, analyses based on these variables do not necessarily maximize the chances of finding task \times difficulty interactions. Also, Ashcraft et al. did not include a detailed comparison of errors in production and true verification. This is not particularly surprising because, until recently, RT analyses have been the mainstay of modern research on the simple number combinations. In verification, however, whether an error is due to an incorrect retrieval, an error at the decision stage, or an error at the response stage, is never clear. Given the additional ways in which errors can occur in verification, it may be that the comparability of production and true verification is restricted to RT data.

One purpose of Experiment 1 was to conduct a more exhaustive comparison of production and true verification, including detailed analyses of errors. The experiment used single-digit multiplication problems. Data from pilot studies were used to estimate the normative difficulty of the problems. Thus, for this study, problem difficulty was defined empirically, which may enhance the ability to detect task \times difficulty interactions. Experiment 1 also investigated a false-answer effect that occurs in verification of multiplication: the *confusion-product* effect (Stazyk et al., 1982). This phenomenon has implications both for the nature of the representation underlying multiplication skill and for the comparability of verification and production. Specifically, the effect may indicate that presenting certain false answers can disrupt retrieval by increasing associative interference. If this is so, then the presented answer does have an effect on arithmetic retrieval processes, and the conditions under which, and the extent to which, the verification task can provide an accurate picture of these processes becomes an empirical question. Before discuss-

ing the confusion-product effect, the evidence for interference in mental arithmetic should be considered.

Interference in the Production of Arithmetic Facts

Several kinds of evidence suggest that arithmetic retrieval errors can be understood as intrusions by false associations and that correct RT is influenced by the strengths of false associations (Campbell, 1985, 1987; Campbell & Graham, 1985; Siegler, 1986a; Siegler & Shrager, 1984). Campbell and Graham (1985) analyzed errors made by children and adults in single-digit multiplication (see also Graham, 1987; Miller et al., 1984; Norem & Knight, 1930) and found that a high percentage of errors (80% in adults) were correct answers to other problems in the same times-table (so-called *table-related* errors; e.g., $3 \times 7 = 24$ or $3 \times 7 = 28$). Campbell (1985) found a comparable pattern in the errors made by adults on simple division problems. When a dividend was evenly divisible by more than two single-digit numbers (e.g., 12, 16, 24, and 36), 82% of errors involved an answer that was correct to another problem involving that dividend (e.g., $24/3 = 6$). Thus, division errors also tend to be table related.

Campbell and Graham (1985) proposed an associative-interference account of table-related errors in multiplication. They suggested that associations between multipliers and products acquired in the context of one problem (e.g., learning $4 \times 7 = 28$ involves associating each of 4, \times , and 7 to 28) are activated, and sometimes retrieved, in the context of other problems involving those multipliers (e.g., $4 \times 6 = 28$, $3 \times 7 = 28$). Over time, this produces permanent associative links between problems and false products in the same times-table. Campbell and Graham suggested that the error patterns reflect an associative-network structure in which each problem may be linked to several candidate answers and each product linked to a number of different problems. In several essential respects, the ideas proposed by Campbell and Graham are similar to the distribution-of-associations model offered by Siegler (1986a; Siegler & Shrager, 1984). Encoding a problem activates a set of response candidates, and the level of activation of each candidate is proportional to its acquired strength (cf. Anderson, 1981, 1983). It is assumed also that the probability of a candidate's retrieval is related directly to its level of activation. Consequently, the frequency of a specific error is taken as an estimate of associative strength.

In both the network-interference account and the distribution-of-associations model, the incidental activation of false associations interferes directly with the speed of correct retrieval (see also Anderson, 1983, for a discussion of "fan" effects). One source of evidence for this is the observation that error rates and correct RT are positively correlated across problems. Siegler (1986a) found a correlation of .87 between RT and error rates for third-graders' multiplication. Campbell and Graham (1985)

reported correlations ranging from .69 to .81 for children's multiplication in Grades 2 through 5 and a correlation of .85 for university students. Testing adults, Matthews (1985) found a correlation of .74 between problem error rate and correct RT in simple addition. Campbell (1985) reported a correlation of .73 between RT and errors in simple mental division by adults. Thus, across simple arithmetic problems, correct RT is directly proportional to the probability of an error. Taken together with the associative nature of errors, the strong correlations between errors and correct RT across problems suggest that RT is slowed by interference generated by false associations.

Campbell (1987) provided more direct evidence for an interpretation of errors as associative intrusions, and for the hypothesis that the strengths of false associations (as estimated by the frequencies of specific errors) influence the speed of a correct retrieval. Testing adults on multiplication problems, Campbell showed that retrieval of a product via one problem (e.g., 24 via 3×8), can slow correct RT or promote an error on subsequent problems (e.g., 4×8). This occurs, however, only when the latter have relatively high normative probabilities of generating that product as an error. Hence, the effect is referred to as *error priming*. Problems are minimally affected or totally unaffected by prior retrieval of products that are infrequent error responses. Campbell argued that the error-priming effect occurs because retrieval of an answer primes the corresponding node in the network. This produces residual activation that temporarily amplifies interference for other problems that possess a false associative link to that node. The proposed locus of interference is a retrieval process that functions to provide the "best" candidate response, probably on the basis of relative activation levels (cf. Anderson, 1981, 1983; Gillund & Shiffrin, 1984; Rundus, 1973; Siegler, 1986a, 1986b). A manipulation that increases the activation of a node relative to the activation levels of competing nodes increases the probability that this node will be selected and decreases the selection time. A manipulation that decreases relative activation has the opposite effects. Consequently, increased activation of a false candidate slows down retrieval of the correct answer.

The interfering effects of error priming on RT indicate that there is a direct relationship between the speed of correct retrieval and the strengths of false associations, where strength is measured by the frequency of a particular error. Thus, Campbell's (1987) results support the hypothesis that the strong positive correlations observed between error rate and correct RT across problems reflect an associative-interference process that influences both (Campbell & Graham, 1985; Graham, 1987; Siegler, 1986a, 1986b; Siegler & Shrager, 1984).

Retrieval Interference and the Confusion-Product Effect in Verification

Using the verification task, Stazyk et al. (1982) found a multiplication phenomenon that may be directly related

to the error-priming effect reported by Campbell (1987). They referred to this finding as the confusion-product effect. RT for a false answer that was a table-related product was over 100 msec slower than RT for a false answer that was an unrelated number. Since table-related products are the answers most frequently generated as errors (i.e., they have strong false associations), the confusion-product effect might involve the same mechanisms responsible for the error-priming effect. In verification, the presented answer may prime (i.e., boost the activation level of) its corresponding node in the network (Ashcraft, 1986), much in the same manner that Campbell (1987) argued that retrieval of an answer results in priming. If the presented problem possesses a false associative link to this node, activation from priming and from the problem may combine to disrupt retrieval by increasing the total amount of false activation relative to the activation of the correct node. If this account is correct, the confusion-product effect in verification occurs during retrieval and represents converging evidence for the network-interference theory.

Although Ashcraft, Koshmider, Roemer, and Faust (1985; see also Stazyk et al., 1982) acknowledged that a presented answer may affect retrieval performance, in Ashcraft's (1986) network-retrieval model the slowing effect of presenting table-related answers in verification is attributed to strength-related interference at the post-retrieval comparison-decision stage. When the correct answer and the presented false answer are similar in associative strength, the comparison-decision operations take longer than when the strengths are dissimilar. Attribution of the effect to the decision stage is consistent with the assumption that the presented answer in verification has no effect on the duration of the retrieval stage (Ashcraft et al., 1984; Parkman & Groen, 1971). If presenting false answers in verification can induce interference at retrieval, however, it follows that the retrieval processes in verification and production are subject to different influences. Furthermore, if priming false answers interferes with retrieval then, conversely, presenting the correct answer may facilitate performance (i.e., facilitate the retrieval process by increasing the relative activation level of the correct answer). Depending on the nature of such an effect (e.g., whether it is uniform across problems or interacts with problem difficulty), the validity of generalizing true verification performance to the production task becomes uncertain. Thus, the locus of the confusion effect has direct implications for understanding the possible effects on retrieval of the presented answer in verification.

Experiment 1 attempted to replicate the confusion-product effect and to investigate that effect's relation to the frequency of errors in production. It is not clear that the effect reported by Stazyk et al. (1982) is related to the associative strength linking problems to specific false answers (Baroody, 1984). The non-confusion products they used not only were multiplicatively unrelated to problems, but at least some of them also were not familiar

multiplication answers at all (e.g., $2 \times 7 = 19$). Subjects might have rapidly identified some unrelated answers as implausible candidates, perhaps eliminating the multiplication component of the task. Thus, one purpose of the present experiment was to test the confusion effect attributable to table-related products (e.g., $3 \times 4 = 15$) against a control comprising table-unrelated products (e.g., $3 \times 4 = 14$). This condition constitutes a replication of the confusion effect reported by Stazyk et al. (1982), but under conditions in which the unrelated answer is always plausible (i.e., it is a familiar correct answer to some single-digit multiplication problem).

Theoretically, the two conditions represent a manipulation of strengths of false answers. If error rates index associative strength, then problems tend to have stronger false associations with table-related products than with table-unrelated products. As a check that the functional difference between these conditions is tied to the relative frequencies of specific errors (i.e., to associative strength), a second false condition was included. For a subset of problems, one false answer assigned to each problem was table related, but was an infrequent error response. The other was the most frequent error response to each problem, as determined in pilot studies.¹ If the confusion effect depends on the strength of the link between a problem and the presented false answer, the effect should be greater when a false answer is a common-error response than when false answer has a low probability of being generated as an error, even when the latter is table related.

In summary, there were two purposes for Experiment 1. One was to compare multiplication performance in the production task with performance in true verification, with the specific purpose of determining whether the relative difficulty of problems is the same in the two tasks. The second purpose was to replicate the confusion-product effect and to test whether the interfering effect of presenting table-related products in false verification is associated with the probability of generating that product as an error.

EXPERIMENT 1

Method

Subjects

Nine female and 7 male undergraduate psychology students from the University of Waterloo participated as part of a course requirement. The subjects were 19 to 22 years old, and all reported normal or corrected-to-normal vision.

Apparatus

For both the production and the verification tasks, the problems were presented on a black-and-white television monitor connected to a PET 2001 series microcomputer. The Commodore Business Machine standard character set (about 5 mm high \times 3 mm wide) was used, and the stimuli appeared as white characters on a dark background. RTs were determined with a software clock accurate to ± 1 msec. Timing began with the onset of the stimulus. In the verification task, the timer stopped when subjects pressed one of two buttons signifying true or false. In the production task, the timer stopped when a voice-activated relay fed by a lapel microphone worn by the subjects detected the subjects' verbal responses.

Stimuli and Design

Each subject was tested in one session lasting approximately 2 h and involving a set of production trials and a set of verification trials. There was a break of about 10 min between tasks. Task order and the assignment of left and right buttons to true and false responses were completely counterbalanced over subjects.

In the verification task, the subjects made true-false decisions about arithmetic equations of the form $a \times b = c$, with the actual multiplication problems ranging from 2×2 to 9×9 . Presented horizontally at the center of the screen, the problems occupied a 9- or 10-character field, depending on whether the presented answer was a one- or two-digit number. If operand order is disregarded (i.e., if $3 \times 4 = 12$ and $4 \times 3 = 12$ are treated as the same problem), there were 36 different problems: 8 ties (e.g., 4×4 , 6×6 , etc.) and 28 nonties.

One purpose of Experiment 1 was to look for evidence of an interaction between task (production vs. true verification) and problem difficulty. On the basis of problem error rates ascertained from the data of pilot experiments (see Note 1), the 36 problems were divided into two sets. There were 16 high-error problems in one set and 20 low-error problems in the other (see Appendix A). The normative multiplication data presented by Campbell and Graham (1985, Appendix B) confirm this breakdown. Those data indicate an overall error rate of 3.2% for the 20 problems assigned to the low-error set and a 13.3% rate for the 16 high-error problems. As would be expected given the overall relation between problem size and difficulty, the low-error set comprises mainly small-number problems and the high-error set mainly large-number problems. "Tie" problems (e.g., 6×6 , 8×8 , etc.), which are consistently found to be among the easiest problems (e.g., Miller et al., 1984), all appear in the low-error set.

Each problem was tested twice with the correct product. In addition, two different false answers were assigned to each problem, one a strong false associate and the other a weak false associate. For the low-error set, one false answer was multiplicatively related (i.e., a table-related product), whereas the other false answer was also a product in the range of multiplication problems tested but was not a multiple of either of the problem's multipliers (i.e., table-unrelated).² Although it is true for the low-error problems that table-related errors are more likely to occur than table-unrelated errors, it is unusual for any specific error (e.g., $2 \times 5 = 15$) to be particularly frequent (see Campbell & Graham, 1985). For the more error-prone problems, however, it is often possible to identify for individual problems a specific answer that occurs commonly as an error and a table-related answer that is an uncommon error response. This provides an opportunity to test specifically whether the confusion-product effect varies with the strength of a false association, where strength is estimated by the normative frequency of an error response. Thus, for each problem in the high-error set, one false answer was a table-related product, but one that does not occur frequently as an error to that problem: the low-frequency (LF) table-related error condition. The other was the most frequent error for that problem, according to errors observed in the pilot experiments. All but one of the latter ($6 \times 7 = 32$) were table-related products in addition to being frequent error responses. This will be referred to as the high-frequency (HF) error condition. For 7×9 , for example, the LF table-related answer was 49, while 56 was the HF-error answer. In the table of multiplication errors presented by Campbell and Graham (1985, Appendix A), the particular false relationships (e.g., $7 \times 9 = 56$) selected for the HF-error condition accounted for 33.2% of all errors made on the problems in the high-error set, whereas the LF-error answers accounted for only 8.0%. In addition to the breakdown of low-error and high-error problems, Appendix A also presents the false answers assigned to each problem.

The average difference between false answers and correct products (the split) and the frequencies of positive and negative splits were balanced across conditions for the low-error and high-error sets.

For the 20 low-error problems, the average absolute split for false table-related answers was 4.85, with 10 of 20 being less than the true answer. The average split for the 20 table-unrelated answers was 4.05, and again 10 of 20 were less than the correct product. For the 16 high-error problems, the mean absolute split for LF table-related answers was 8.75, with 13 of 16 being less than the true value. The average absolute split for the HF-error answers was 7.81, with 14 being less than the correct answer. Although the average splits for the two false conditions within each problem set were not identical, the remaining differences in split (an average difference < 1 for each problem set) are substantially smaller than those that have been found to produce significant effects in false verification using multiplication stimuli (see Stazyk et al., 1982, Experiment 1).³

Each problem was presented once in each of four blocks of 36 verification trials. In each block, there were equal numbers of true and false trials, and trial type was randomized with the constraint that no more than three trials of the same type occur consecutively. In addition, there was no repetition of either of the multipliers or of a correct or presented answer on consecutive trials.

Across blocks, each problem appeared twice in a true equation and twice in a false equation. In each of the four blocks, 10 of the 20 low-error problems were tested under the true condition. Five of the remaining low-error problems were tested with false table-related answers, and the other five with table-unrelated products. For the 16 high-error items, 8 were tested with the correct product in each block, 4 were tested with the HF error products, and 4 were tested with the LF-error products.

For each subject, problems were assigned randomly to an order of conditions across blocks. Given that each problem is tested with the correct answer twice, and once in each of two false conditions, there are 12 unique orders of conditions across four blocks of trials. There were 20 orders determined for the low-error problems and 16 for the high-error problems. For each problem set, the orders of conditions included the 12 unique permutations. The remaining orders for each set were arbitrarily chosen to preserve the distribution of conditions within and across blocks described above.

Across blocks, true nonties were tested once with the numerically smaller multiplier on the left and once with the larger on the left. False nonties were also tested once in each orientation, randomly with respect to false-answer type. Within each block, half of the true trials and half of the false trials for each problem set were tested in one orientation, and the other halves tested in the reverse orientation.

For the production task, the same 36 problems served as stimuli. The problems appeared in a five-character field: the two multipliers separated by an uppercase X with adjacent blank characters. Each problem was tested once in each of four blocks of 36 trials. Trial order was randomized independently for each block for each subject, with the constraints that consecutive problems did not share a common multiplier nor a common product. For each subject, 14 of the 28 nontie problems were assigned randomly to have the smaller multiplier appear on the left in the first block. Operand order for each problem then alternated across blocks so that each nontie problem was tested twice with each order.

Procedure

The subjects were tested individually; during testing, they sat about 50 cm from a monitor located in a quiet, slightly darkened room.

For the verification task, the subjects were informed about the range of multiplication problems they would encounter and were told that trial presentation would be continuous in each of four blocks of 36 problems. In addition, they were instructed to emphasize both speed and accuracy and were advised that true and false trials were equally likely. Each trial block was initiated by the experimenter, who remained present throughout testing. On each trial, a fixation dot appeared at the center of the screen for 3 sec and then flashed

off and on twice over a 1.5 sec interval. The problem appeared on what would have been the third flash. The rightmost multiplier occupied the position of the fixation point. The buttonpress response cleared the screen, and the fixation dot for the next trial appeared immediately. No feedback was provided to subjects about the accuracy of their responses. The response and RT for each trial were recorded by the computer. About 15 sec separated trial blocks in the verification task.

In the production condition, the subjects were required to state the answer to a presented problem. They were informed about the range of problems to be tested and were encouraged to stress both speed and accuracy. The subjects were familiarized with the operation of the voice-activated relay and were instructed to avoid making extraneous sounds that would stop the timer prematurely. Triggering the relay caused the problem to disappear immediately from the screen, which allowed the experimenter to detect and record failures of the voice key. The experimenter initiated each block of trials. On each trial, a fixation dot appeared briefly at the center of the screen and then flashed twice over 1.5 sec. The problem appeared as the third flash with the X appearing at fixation. Following each response, the experimenter immediately entered the given answer at the computer keyboard. Provided that this was accomplished within 3 sec, a constant 4.5-sec interval separated the end of one trial from the beginning of the next. Regardless of the duration of response entry, the subjects always received the two warning flashes before a problem appeared. About 15 sec separated blocks of trials.

Results and Discussion

Seventy-two production RTs (3.7%) were spoiled by failures of the voice key. Additionally, production RTs were discarded as extreme scores if they were more than 2.5 standard deviations (*SDs*) from a subject's grand mean for correct RTs. There were 69 RTs identified as outliers (3.1% of the 2,232 unspoiled trials). The overall error rate for production trials was 6.0% (139 errors). For verification, a grand mean and *SD* for correct true trials and a grand mean and *SD* for correct false trials were computed for each subject. Extreme RTs were identified separately for true and false trials on the basis of the outlier rule described above. Of the 1,152 true trials, 26 (2.3%) were discarded, and 30 (2.6%) of the 1,152 false trials were excluded as outliers. The subjects made a total of 88 errors on true trials (7.6%) and 77 errors on false trials (6.7%).

Comparison of Multiplication Production and True Verification

As pointed out earlier, a proper comparison of production and verification is restricted to the use of true verification trials, because it is known that there are unique effects in false verification. If, however, true verification simply adds a comparison-decision stage that affects each trial equally, production and true verification should reveal similar patterns of RTs and errors across problems.

Error data. The error rates under production and true verification for the low-error and high-error sets were computed for each subject. Table 1 presents the percentage of errors in each condition. Errors were analyzed in a repeated measures analysis of variance (ANOVA) with factors of task (production vs. true verification) and

Table 1
Mean Correct Response Time (RT, in Milliseconds) and Percent Errors (PE) from Production and True Verification Tasks for the Low-Error and High-Error Problem Sets

	RT	SD	PE	SD
Production Task				
Low Error	798	144	2.9	3.2
High Error	998	275	10.0	7.2
True Verification Task				
Low Error	928	199	7.0	8.1
High Error	1094	279	8.4	9.3

problem set (low-error vs. high-error problems).⁴ There was no main effect of task [$F(1,15) < 1$], but the problems assigned a priori to the high-error condition did indeed yield more errors than the low-error items [$F(1,15) = 12.09, p < .005$]. The interaction of task and problem set was highly significant [$F(1,15) = 8.83, p < .01$]. The interaction reflects the fact that the high-error set was substantially more error prone than the low-error set when tested using production [a difference of 7.1%, $t(15) = 4.75, p < .001$], but not when tested using verification [a 1.4% difference, $t(15) < 1$]. Thus, whereas the prior assignment of problems to the low-error and high-error sets (on the basis of pilot studies and verified by normative data) was confirmed by the production task, this distinction was not preserved in true verification performance. Across the 36 problems, however, the number of errors made in production and true verification correlated reliably [$r(34) = .43, p < .01$]. This relationship, which captures only 18% of the variance in errors, seems modest when one considers that the correlation involves a pairwise comparison of accuracy for the same problems.

Response time data. For each subject, mean RTs in the production task for the low-error and high-error problems were computed by taking the median of correct RTs for each problem (a maximum of four) and then computing the means of the problem medians (see Miller et al., 1984). For each subject, the average of correct true-verification RTs for each problem (a maximum of two) was computed, and the mean of the averages was taken across the problems. Because an individual's RTs can vary greatly for different problems, a mutual-exclusion rule for problem RTs was applied across conditions within subjects. If no RT for a problem was available in one task because of spoiled trials, outliers, or errors, the RT for that problem was excluded in the other task before a subject's means were calculated. This ensured that the same problems contributed to each subject's means for each task. After applying mutual exclusion, an average of 19.7 of a maximum of 20 problems contributed to subjects' means for the low-error set. For the high-error set, the average was 14.8 of a maximum of 16 problems. Table 1 includes the means of the subjects' means and the standard deviations.

A 2×2 (task \times problem set) repeated measures ANOVA indicated that the high-error set yielded longer

RTs than did the low-error set [$F(1,15) = 27.21, p < .001$] and that true verification required longer RTs than did production [$F(1,15) = 25.92, p < .001$]. The mean difference between the low-error and high-error problem RTs was 34 msec greater in production than in true verification, which suggests the same trend as in the error data, but the task \times problem set interaction for RT was not reliable [$F(1,15) = 2.95, p > .05$].

Mean correct RTs were computed for each problem, and the pattern of RTs across problems in each task was subjected to a problem-size analysis, according to the general procedure used by Ashcraft et al. (1984) to compare verification and production RTs for mental addition. All correct problem RTs, except those discarded by the original outlier rule, contributed to the mean for each problem. For each task, RTs across the set of problems were correlated with the standard problem-size variables (the minimum or maximum multiplier, the sum or squared sum of the multipliers, or the correct product; see Miller et al., 1984; Stazyk et al., 1982). The value of the numerically larger multiplier (max) yielded the highest correlation among the five problem-size variables for both production [$r(34) = .68, p < .001$] and true verification [$r(34) = .58, p < .001$].⁵ The equations were $RT_v = 787 + \max(32.4)$ for true verification and $RT_p = 617 + \max(42.1)$ for production. With max ranging from 2 to 9, as in the present experiment, the predicted range for verification RTs is 227 msec, whereas for production it is 294 msec. Thus, the analysis by problem-size variables predicts a 29.5% larger range for production RTs than for true verification RTs. As the regression equations suggest, the difference between true verification and production RTs decreased as max increased [$r(34) = -.39, p < .02, b = -16.3$ msec]. Although there was a strong correlation between production and true verification across the 36 item means [$r(34) = .78, p < .001$], the RT advantage for production over true verification tended to be smaller for the more difficult large-number problems than for the relatively easier small-number problems.

As a check on this, the 18 fastest and the 18 slowest problems were identified on the basis of the production-task means. The corresponding true-verification RTs then were used to compute difference scores for each problem. Fifteen of the 18 fastest problems in production were also among the 18 fastest problems in true verification. The mean difference between production and true verification RTs for the 18 slow problems was 55.2 msec ($SD = 83.2$), in favor of production. The comparison between the 18 fastest problems in production and the corresponding true verification RTs indicated a 144.2 msec ($SD = 68.4$) advantage for production. A t test for uncorrelated samples indicated that this 89-msec difference was highly reliable [$t(34) = 3.51, p < .01$].

With respect to both accuracy and speed of correct retrieval, true verification seems to attenuate differences across problems that are observed in production. This extends to the relationship between errors and correct RT.

In these production data, the correlation between number of errors and mean correct RT across problems was $r(34) = .75$ ($p < .001$), whereas in true verification, the correlation was $r(34) = .28$ ($p > .05$). This important relationship, observed strongly in production, was not found in true verification.

Discussion of Production and True Verification

Ashcraft et al. (1984) argued that there are no major differences in results derived from using production and verification to study mental addition. The present data do not support an extension of this conclusion to mental multiplication. An important component in the argument for the comparability of true verification and production offered by Ashcraft et al. is that the presented answer affects the comparison-decision stage, but has no effect on the performance of the arithmetic retrieval stage that is common to true verification and production. Effects due to the comparison stage in true verification ought to be constant across problems (e.g., add a constant to the RT or error rate for each problem). It was found here, however, that with the production task there were significant differences in error rates for the low-error and high-error problem sets, whereas the verification task indicated no differences in error rates between these two sets of problems. Although a comparison of RTs using these particular sets did not find a significant interaction with task, the item analyses indicated a greater difference between RTs for problems with fast times in production and their corresponding verification times than between problems with slow times in production and their true verification RTs. Thus, differences between the two tasks as a function of problem difficulty are manifested both in accuracy and speed.

The differences in accuracy may reflect the unique contribution in verification of errors from the comparison-decision and true-false response-selection stages. If a significant percentage of verification errors originate at these points, then retrieval error rates for each problem would be masked. This could explain the relatively low correlation for error rates between true verification and production and the absence of a significant correlation between errors and correct RT across problems in true verification. As explained earlier, there are good reasons to believe that the relation between errors and correct RT reflects response competition during retrieval (Campbell, 1987), and one should expect to find this correlation strongest when problem error rates include retrieval errors only. The results suggest that the verification task may be inherently limited in its capacity to measure retrieval accuracy.

The pattern of errors across conditions suggests that there is more to this limitation than uncertainty about the source of errors in verification. There was a 7% difference between error rates for the low-error and high-error sets in the production task (10.0% vs. 2.9%). In order for a difference of this magnitude to be masked by post-retrieval errors in verification, a much higher rate of er-

rors than was observed would be required. Whereas the error rate for low-error verification problems was more than twice that for production problems (7.0% vs. 2.9%), the rate for the high-error set was actually higher for production than for verification (10.0% vs. 8.4%). The difference between production and verification for the low-error set suggests that a substantial percentage of verification errors are postretrieval, because it is unlikely that presenting the correct answer in verification would induce retrieval errors. If this is correct, however, then the crossover in error rates for the high-error problems indicates that presenting the correct answer in verification can reduce sharply the probability of a retrieval error. One possibility is that presentation of the answer activates or primes that response in the multiplication retrieval structure. Priming, however, may interact with accessibility. Correct answers that are activated to a high level directly by their problems may receive relatively little benefit, whereas the accessibility of weakly activated products may be enhanced significantly by additional activation from priming.

The pattern of RTs is consistent with this account. Although, overall, true verification required longer RTs than did production, there were smaller differences between problems with slow RTs in production and their corresponding verification RTs than there were for problems with fast production RTs. This suggests that the comparison-decision stage increases RT overall, but that retrieval times for problems with less accessible correct products are speeded more by priming the correct answer than are retrievals for problems with highly accessible products. Experiment 2 directly tested the hypothesis that priming interacts with the difficulty of generating an answer.

Analysis of the Confusion-Product Effect

Table 2 presents the mean RTs and error rates for the two false conditions for each of the low-error and high-error sets. The low-error and high-error sets correspond to different tests of the confusion-product effect and consequently received separate analyses.

The table-related and table-unrelated false answers assigned to each low-error problem constitute a replication of the manipulation used by Stazyk et al. (1982), but with unrelated products that are all plausible candidate answers. A maximum of 20 problem RTs could contribute to each

Table 2
Mean Correct Response Time (RT, in Milliseconds) and Percent Errors (PE) for the False Verification Conditions in Experiment 1

Condition	RT	SD	PE	SD
Low-Error Problems				
Unrelated Answer	1031	240	3.8	5.3
Related Answer	1122	280	5.6	6.3
High-Error Problems				
LF-Error Answer	1161	312	3.9	3.9
HF-Error Answer	1191	276	14.5	11.6

subject's means for the low-error problems. After discarding outliers and error RTs and applying mutual exclusion, an average of 17.7 problems contributed to subjects' mean RTs. The 91-msec disadvantage for the related condition compared with the unrelated condition was highly significant [$t(15) = 4.07, p < .01$]. There was no comparable effect on errors, however [$t(15) = 1.00$].

The same comparisons for the high-error problem set tested whether the different effects of presenting table-related and table-unrelated answers can be traced directly to the strengths of false associations, where strength is defined by the frequency of the false answer in production. It is important first to confirm that the HF-error false answers assigned to these problems were in fact generated as error responses in this experiment more frequently than were the LF table-related answers. The particular false relationships identified a priori for the HF-error condition (e.g., $3 \times 9 = 18$) accounted for 30.4% of all errors made on the 16 high-error problems in production, whereas the LF-error answers accounted for only 2.9%. There was a potential maximum of 16 problem RTs available for subjects' means in each of the two false high-error conditions. After applying mutual exclusions, 11.4 problems contributed to each subject's means. The 30-msec disadvantage for the HF-error condition relative to the LF condition was not reliable [$t(15) < 1$]. The latter result was qualified, however, by a comparison of error rates in the two conditions. The rate for HF-error answers was 10.6% higher than that for the LF-error answers. The effect on errors was highly significant [$t(15) = 3.39, p < .01$].

The finding that table-related products slowed RT more than table-unrelated products (low-error problem set) effectively replicates the Stazyk et al. (1982) results and verifies that their finding was not attributable simply to differences in the general plausibility of false answers across conditions. Instead, the results suggest that the effect is related to the associative strength of a false answer: The products that occur more frequently as errors in production (i.e., table-related answers) take longer to disconfirm than do the products that occur as errors less often (i.e., table-unrelated answers). The false conditions for the high-error problem set tested more directly whether the interference with table-related products was tied to the frequency of specific errors in production. The fact that HF-error answers were identified erroneously as true more than 3.5 times more often than their LF-error table-related counterparts suggests that a common mechanism underlies the production of errors and the confusion-product effect in verification.

The failure to find an effect on RT for the high-error problems may be attributed to a speed-accuracy trade-off (see Corbett & Wickelgren, 1978; Wickelgren, 1977). Undoubtedly, accuracy on verification would increase if subjects delayed their responses. The 14.5% error rate in the HF-error condition (compared with 3.9% in the LF-error condition) suggests that subjects were unwilling to delay their responses long enough to maintain even

moderately high accuracy under the interference induced by the HF-error answers. According to this argument, if accuracy had been held more constant across false conditions, the effects of interference due to the strengths of the HF-error answers would have emerged clearly on RT for the high-error problems, as it did for the low-error set.

Conclusions From Experiment 1

The comparison of production and true verification indicated that a number of multiplication phenomena observed across problems in production tend to be weakened or masked in true verification. The results challenge the generality of the claim (Ashcraft et al., 1984) that the major difference between production and true verification tasks is a comparison-decision stage in verification that affects performance on each problem uniformly. The interaction of task with problem difficulty raises the possibility not only that the verification task masks retrieval phenomena, but also that the presented answer in verification may have a direct impact on the retrieval process. Specifically, it was suggested that the presented answer functions as a priming stimulus and that priming the correct answer has a larger effect on difficult problems than on easier problems.

Experiment 2 used the production task to examine the priming of correct answers. In one condition, the subjects received a 300-msec preview of the correct answer before a problem was presented, whereas, in another condition, the prime was a neutral stimulus (##). The neutral condition was analogous to the production task in Experiment 1 and provided a baseline for assessing differential effects of priming the correct answer as a function of problem difficulty. The correct-prime condition simulates true verification but eliminates any effects due to the comparison-decision stage in verification. Thus, Experiment 2 permitted a direct assessment of the impact of priming the correct answer on the process of generating an answer. If the neutral and correct prime conditions reproduce the interactions with problem difficulty observed across production and true verification in Experiment 1, it will confirm that the presented answer in verification functions as a priming stimulus and that priming can affect the answer-generation stage.

Experiment 2 also investigated the locus of the confusion-product effect. Experiment 1 demonstrated that the effect is related to the frequency with which a particular false answer occurs in production. This supports Ashcraft's (1986) suggestion that the effect is governed by the associative strength of a false answer. In Ashcraft's model, however, the effect is placed at the comparison-decision stage, an assumption consistent with the view that the presented answer in verification has no effect on performance in the retrieval stage. The error-priming effect in multiplication production (Campbell, 1987) suggests the possibility, however, that the effect occurs during retrieval. In this case, the effect would be analogous, but opposite in direction, to the proposed facilitating effects of priming the correct answer.

Experiment 2 included two false-prime conditions. In one, the false prime was a HF-error response (the "related" condition) to the following problem, whereas, in the other, the prime was a LF-response (the "unrelated" condition). If priming a related false answer interferes with generating the correct answer, then the related condition should yield slower performance than the unrelated condition. Such a result would suggest that confusion effects in verification can arise during retrieval (Ashcraft et al., 1985). If no differences between the related and unrelated conditions are observed, an account attributing the confusion effect in verification to the comparison-decision stage is supported.

EXPERIMENT 2

Method

Subjects

Ten male and 8 female undergraduates from Carnegie-Mellon University participated as part of a course requirement. All subjects reported normal or corrected-to-normal vision.

Apparatus and Stimuli

The apparatus was identical to that used for the production task in Experiment 1. The stimuli included the multiplication problems ranging from 2×2 to 9×9 . If operand order is disregarded, there are 36 problems in this range. The set of 36 problems was divided into a set of 18 "easy" problems and 18 "difficult" problems according to Campbell and Graham's (1985, Appendix B) normative production data for adults. The 18 problems with the fastest mean RTs were the easy problems. Because of the high correlation between error rates and correct RT across problems, this set also included the 17 problems yielding the lowest error rates for that sample.

For each problem, two false answers were selected from the table of frequencies of multiplication errors presented by Campbell and Graham (1985, Appendix A). One false product assigned to each problem was a table-related, common error response to that problem (the related condition), whereas the other false answer was a table-unrelated, uncommon error response (the unrelated condition).⁶ For each problem, the two false answers had the same digit in the tens place (i.e., fell within the same decade). This controlled for possible effects of featural similarities between a prime and the correct answer. For 17 of the problems, the two false answers fell in the same decade as the correct answer. Appendix B lists the easy and difficult problem sets and the false answers assigned to each problem.

Design and Procedure

Each problem was tested once in each of four blocks of 36 trials. Across blocks, each problem was tested under four priming conditions. In one condition, the "correct" product served as the prime. A second condition involved a "neutral" prime composed of two adjacent number signs (##). The related and unrelated false primes defined the other two conditions. There were nine trials with each prime type in each block. Thirty-six orders of the four conditions were constructed, and for each subject problems were assigned randomly to an order of conditions across blocks. Thus, the testing of any particular easy or difficult problem in any prime condition was random with respect to block. The 36 orders were selected so that, among them, all pairwise orders of conditions occurred equally often. For each subject, the order of problems in each block was randomized, with the constraint that neither of the operands, nor the correct answer or the prime, be repeated across successive trials. The order of operands for nontie problems was determined as in

the production task in Experiment 1 and was random with respect to prime condition.

Prior to the multiplication trials, the subjects performed a number-naming task. The latter accustomed the subjects to responding quickly and to the sensitivity of the voicekey. The digits 0 through 9 were presented one at a time in random order in six continuous blocks of 10 trials. A fixation dot appeared at the center of the screen and flashed twice over a 1.5-sec interval. The digit appeared on what would have been the third flash. The subjects were instructed to name each digit as quickly as possible. The verbal response triggered the voice-activated relay and caused the number to disappear from the screen immediately and the fixation dot for the next trial to appear.

Instructions for the multiplication trials described the range of problems that would be encountered and stressed the importance of both speed and accuracy. The subjects were informed that either a numerical or nonnumerical stimulus would flash briefly on the screen at the fixation point before each problem appeared. They also were told that sometimes a numerical stimulus would be the correct answer to the upcoming problem, but that this would occur at random and could not be used to predict the problem. The subjects were told that, although it was important to focus on the fixation dot at the beginning of each trial, their only responsibility was to provide the answer to each problem and that they could ignore the prime.

The experimenter initiated each block of trials. On each trial, a fixation dot appeared at the center of the screen and flashed twice over a 1.5-sec interval. The prime appeared as the third flash, with the rightmost character at fixation. The prime remained on the screen for 300 msec, and then a problem appeared immediately, with the multiplication sign at fixation. The display configuration for problems was identical to the production task in Experiment 1. RTs were measured to ± 1 msec from the onset of the problem until a verbal response triggered the voicekey. All other details of the procedure were identical to the production task in Experiment 1. Testing each subject required about 40 min.

Results and Discussion

RTs were spoiled by failures of the voicekey on 1.6% of trials. Outliers were identified as in Experiment 1, and 2.6% of unspoiled RTs were discarded as outliers. There were 210 errors on multiplication trials (8.1%).

A mean of the correct RTs was computed for each of the four priming conditions for each subject. The mean RT and error rate for each condition are presented in Table 3. A repeated measures ANOVA of RT⁷ indicated a large effect of priming condition [$F(3,51) = 52.35, p < .001$]. A Newman-Keuls test performed on the four means showed that only the difference between the neutral and unrelated conditions did not reach significance at the .05 level. The numbers of errors made in each condition by each subject were analyzed in a one-factor repeated measures ANOVA [$F(3,51) = 12.16, p < .001$]. A Newman-Keuls test indicated that the correct prime condition

Table 3
Mean Correct Response Time (RT, in Milliseconds) and Percent Errors (PE) for the Four Prime Conditions in Experiment 2

Prime Type	RT	SD	PE	SD
Correct	720	156	2.6	2.4
Neutral	826	188	9.3	5.7
Unrelated	845	210	9.6	6.5
Related	872	203	10.9	5.9

yielded significantly fewer errors than any of the other three conditions ($p < .01$ in each case), but that no other differences among conditions were significant. These results confirm that a 300-msec exposure to a numerical prime has a significant effect on speed and accuracy of production of multiplication facts.

Effects of Priming the Correct Answer

Priming the correct answer yielded more than a 100-msec advantage over any other condition and reduced errors by about 70%. The comparison of multiplication and true verification in Experiment 1 suggested that true verification performance may present a distorted picture of the pattern of RTs and errors across problems because of such priming. It was suggested that presenting the correct answer constitutes a priming manipulation, generally facilitating retrieval, but that performance on problems with less accessible answers may benefit more from priming than does performance on problems with correct answers that are already highly accessible.

For each subject, a mean correct RT was computed for the neutral and correct priming conditions for the easy and difficult problem sets. The mutual-exclusion rule was applied across prime conditions for each problem set within subjects. The mean RTs and error rates for the four cells are presented in Table 4.⁸ An average of 16.1 problems contributed to each subject's means for the easy set, and 13.8 problems contributed to each subject's means for the difficult set. A 2×2 (prime condition \times problem set) repeated measures ANOVA indicated that the correct prime condition yielded faster RTs than the neutral condition [$F(1,17) = 84.36, p < .001$] and that easy problems were faster than difficult problems [$F(1,17) = 28.82, p < .001$]. The interaction between these factors was significant [$F(1,17) = 7.80, p < .02$]. The interaction reflects the fact that, relative to the neutral condition, priming the correct answer reduced RT 55 msec more for difficult problems than for easy problems.

An ANOVA on numbers of errors made in each condition showed that the correct prime led to fewer errors than the neutral prime [$F(1,17) = 21.22, p < .001$] and that the difficult problems were more likely to yield errors than the easy problems [$F(1,17) = 8.00, p < .02$]. Although the interaction for errors did not reach significance [$F(1,17) = 3.02, p < .11$], separate dependent t tests indicated that easy problems were less error prone

than difficult problems in the neutral condition [a difference of 8.1%, $t(17) = 2.55, p < .025$], but did not detect differences in error rates across problem sets in the correct condition [a 2.2% difference, $t(17) = 1.44, p > .15$]. Thus, Experiment 2 confirmed that prior exposure of the correct answer facilitates both speed and accuracy of production more for difficult problems than for easier problems.

In Experiment 1, the correlation between error rate and correct RT across problems was robust for the production data but nonsignificant for the true verification data. Similarly, in Experiment 2, priming the correct answer washed out this relationship [$r(34) = .29, p > .05$], whereas it was reliable in the other three conditions ($p < .01$ in each case), averaging .57. This suggests that the failure to find this correlation in true verification is due to a sharp reduction in retrieval errors for more difficult problems and to a flattening of RT across problems that occurs when the correct answer is primed.

The neutral and correct conditions reproduced the pattern of results observed in the comparisons of production and true verification in Experiment 1. Taken together, the results of these experiments suggest that the presented answer in verification may have substantial effects on the retrieval process underlying simple multiplication.

Effects of Priming False Answers

The initial ANOVA showed that RTs in the related condition were longer than those in the unrelated condition and that there was no difference between unrelated and neutral primes. Two further analyses were conducted to check these results. First, the subjects' means were recomputed after the mutual-exclusion rule was applied, which ensured that each pair of means for each subject was based on the same problems. After mutual exclusion was applied, the means were 858 msec ($SD = 195$) for the related condition and 833 msec ($SD = 207$) for the unrelated condition, with an average of 26.6 problems of a maximum of 36 contributing to each subject's means. A dependent t test indicated that the 25-msec difference was reliable [$t(17) = 2.81, p < .02$]. The means for the neutral and unrelated conditions after mutual exclusion were 815 msec ($SD = 179$) and 839 msec ($SD = 208$), respectively (on average, means were based on 27.5 problems). In contrast to the results of the initial ANOVA (no mutual-exclusion rule applied), a dependent t test indicated that the 24-msec advantage for the neutral condition was significant [$t(17) = 2.26, p < .05$].

A more detailed analysis of the false-answer conditions used means computed for each problem in each condition. Multiple regression was used to investigate whether the effects on RT of priming false answers varied as a function of (1) the absolute difference between the correct answer and the prime (split), (2) whether or not the prime was in the same decade as the correct answer (decap), (3) easy and difficult problems (ez/diff), and (4) the related versus unrelated false prime conditions (rel/unr). The coding of these variables is explained in

Table 4
Mean Correct Response Time (RT, in Milliseconds) and Percent Errors (PE) for the Easy and Difficult Problem Sets in the Neutral and Correct Prime Conditions in Experiment 2

	RT	SD	PE	SD
Neutral Prime Condition				
Easy	756	150	5.2	5.5
Difficult	907	242	13.3	11.1
Correct Prime Condition				
Easy	668	124	1.5	3.2
Difficult	764	200	3.7	4.7

Table 5
Correlation Matrix for the Regression Analysis of RTs in
the False-Answer Prime Conditions

	Split	Decrep	EZ/Diff	Rel/Unr
Δ RT	.325	.202	-.082	.277
Split		.553	.072	.027
Decrep			.026	.000
EZ/Diff				.000

Note— Δ RT = a problem's RT in the related and unrelated prime conditions minus its mean RT in the Neutral prime condition. Split = the absolute difference between a problem's correct answer and its false-answer primes. Decrep = 0 when a false prime is in the same decade as the correct answer, and 1 when it is not. EZ/Diff = Easy problems (0); Difficult (1). Rel/Unr = Related prime (1); Unrelated prime (0). With 68 degrees of freedom, a correlation of .24 is significant at $p < .05$.

Table 5. The dependent variable was the difference between the mean RT for a problem in the neutral condition and its mean RTs in the related and unrelated conditions. The problem 7×9 was excluded from the analysis because only 7 of 18 subjects contributed RTs to the mean for 7×9 in the related condition.⁹ Each of the remaining 35 problems contributed two RT difference scores (related—neutral and unrelated—neutral), for a total of 70 cases. Table 5 presents the matrix of zero-order correlations. Independent variables were entered according to a forward stepwise procedure. Split ($b = 2.6$ msec per unit split) and rel/unr ($b = 38.5$ msec) entered the regression on successive steps [$R = .421$; $F(2,67) = 7.22$, $p < .005$], but neither ez/diff nor decrep captured significant variance. A second regression analysis indicated that RT differences between the related and unrelated conditions did not vary as a function of ez/diff [$r(33) = -.16$, $p > .05$].¹⁰

Although the significance of rel/unr simply confirms that the related condition slowed RT relative to the unrelated condition, the additional effect of split is rather surprising. In contrast to the usual effect of split in verification (which is almost certainly an effect associated with the comparison-decision stage; see Ashcraft & Stazyk, 1981), RT here tended to increase with larger differences between the correct answer and the prime. This result may reflect subjects' efforts to take advantage of predictive information provided by the prime. In general, the prime was a good predictor of the magnitude of the correct answer, with splits greater than 10 accounting for less than 15% of the trials on which a numerical prime appeared. Thus, the subjects could usually, but not always, rely on the prime to estimate the numerical range of the answer to the upcoming problem. There is evidence that subjects can take advantage of probabilistic information to facilitate performance in simple arithmetic tasks and that performance on exception trials suffers (Zbrodoff & Logan, 1986, Experiments 3 and 4), but exactly how such an effect would have operated in the present experiment is unclear. Another speculative possibility is that magnitude is a primeable arithmetic dimension: Retrieval performance may be facilitated when evidence about the magnitude of the correct answer is activated.

Discussion of False-Answer Priming

Ashcraft (1986) argued that the confusion-product effect in verification originates at a postretrieval stage in which the retrieved answer is compared with the presented answer. The present results show, however, that processing an associated false answer slows RT even when no explicit comparison stage is involved. The interfering effect of priming false answers appears to be maximal when the prime is a common error response to the upcoming problem. The connection with error frequency suggests that the interference effect in the present experiment may involve the same mechanisms responsible for the error-priming effect reported by Campbell (1987). He found that a retrieval of a product via one problem (e.g., 24 via 3×8) can slow correct RTs to subsequent problems (e.g., 4×8) when the latter have a relatively high normative probability of generating that product as an error. Campbell argued that problems activate multiple responses in an associative-network structure and that retrieval interference is proportional to the relative activation levels of false responses. Retrieval of a response primes the corresponding node in the network, and the additional activation temporarily amplifies interference for other problems that possess a relatively strong false associative link to that node.

By analogy to the error-priming effect, presenting a false answer visually may similarly boost the activation of the corresponding node in the retrieval network. If the upcoming problem possesses a strong associative link to this node, the combined activation from priming and from the problem increases retrieval interference. If the primed node has no link to the problem or only a weak associative link, priming that response has little or no effect on retrieval performance.

GENERAL DISCUSSION

The results of the current experiments challenge theoretical models of arithmetic verification in which the presented answer is assumed to have no effect on the duration of the answer-generation stage (e.g., Ashcraft et al., 1984; Parkman & Groen, 1971). The possibility of such an effect was suggested by the interaction of task (production vs. true verification) and problem difficulty in Experiment 1. Experiment 2 presented to subjects in a production task the same information provided by the presented answer in verification. It was confirmed that presentation of both correct and false answers can affect production performance. The interactions of correct-answer priming with problem difficulty, the suppression of the correlation between errors and RT across problems, and the interfering effects of presenting associated false answers that occur in verification were all obtained in the production version of the task. Taken together, the results of Experiments 1 and 2 indicate that the presented answer in verification is a priming stimulus that can directly affect the process of generating an answer.

The facilitating and inhibiting effects of priming are consistent with both the network-interference (Campbell, 1987; Campbell & Graham, 1985; Graham, 1987) and the distribution-of-associations (Siegler, 1986a; Siegler & Shrager, 1984) models of memory for arithmetic facts. According to these accounts, a problem activates a number of response candidates to various levels of activation, and retrieval performance is determined by the activation level of the correct response relative to the activation level of its competitors (see also Anderson, 1981, 1983). Consequently, a manipulation such as priming, which increases the activation level of a correct answer relative to the levels of its competitors, facilitates correct retrieval. Experiments 1 and 2 demonstrated that priming the correct answer improves retrieval of less accessible answers more than it does retrieval of highly accessible answers. This may indicate simply that performance on some problems is approaching a maximum. For problems with exceptionally strong correct associations and few or weak false associations, the relative strength of the correct answer may be so high that additional activation from priming yields no appreciable effects on performance. Problems possessing correct answers with low relative strength, however, can gain significant improvements in performance as a function of priming.

Conversely, because performance depends on relative activations, the priming of an associated competitor interferes with correct retrieval. In this case, the activation of the correct answer is lowered relative to the cumulative activation of false associations. This effect appears to be related to the frequency with which an answer occurs as an error in production. If we take error rates as an estimate of associative strength, then the strength of the link to a false answer determines the magnitude of interference induced by priming. The retrieval-interference hypothesis provides a single explanation for the interfering effects of prior retrieval of associated false answers (i.e., the error-priming effect; Campbell, 1987), the effects of visually priming table-related answers (Experiment 2), and the confusion-product effect in verification (Experiment 1; Stazyk et al., 1982). The same or similar mechanisms may also be responsible for other false-verification effects, such as the finding that performance is disrupted by false answers that are correct via another operation (e.g., $4 \times 7 = 11$) (Winkelman & Schmidt, 1974; Zbrodoff & Logan, 1986). Miller et al. (1984) found that a substantial proportion of production errors in multiplication and addition involves cross-operation confusion. Such errors provide evidence that there are associative connections between arithmetic operations. In the present framework, this provides a basis for predicting retrieval interference in false verification. All of these phenomena can be explained by assuming that problems activate specific false associations to varying degrees and that retrieval interference varies with the activation level of false candidates. The additional activation from priming, produced either by retrievals or by

visual encoding, temporarily amplifies the effects of interfering associations that are already present.

Although a retrieval locus of priming effects is implicated by Experiments 1 and 2, there may also be other effects induced by presenting numerical primes in arithmetic tasks. The overall advantage for the correct prime condition in Experiment 2 (see Table 3) probably includes effects beyond those attributable to network priming. In particular, presenting the correct answer may directly facilitate the execution of the verbal response as well (e.g., preactivate an articulatory code or facilitate lexical access). Note, however, that response-execution effects do not explain the correct versus neutral \times easy versus difficult interaction or why strongly associated false primes yielded slower production RTs than did unrelated false primes. The latter phenomena suggest specifically that primes affect an associative retrieval process. It is also important to note that although the confusion-product effect can be obtained in the production task (Experiment 2), this does not rule out Ashcraft's (1986) view that there is an effect of related false answers on the comparison-decision stage in verification. The 46-msec difference between the neutral and related prime conditions (see Table 3) is approximately half the size of the confusion effect found by Stazyk et al. (1982) using verification. Thus, it is possible that both retrieval and comparison-decision stages in verification are affected by the presented answer. Finally, LeFevre et al. (1986) reported a visual-matching study in which subjects made yes/no decisions about whether either of the addends in a simple addition problem matched a comparison digit. "No" decisions were slower when the comparison digit was the correct answer to the addition problem than when the comparison item was unrelated to the problem. This suggests that models of response-competition mechanisms developed to explain facilitation and interference effects in visual search and visual matching tasks (e.g., Eriksen & O'Hara, 1982; Proctor, 1981) may also be relevant to our understanding of arithmetic priming effects.

Another issue of Experiments 1 and 2 concerned the ability of the verification task to provide an accurate picture of the processes underlying arithmetic production. Educators are fundamentally interested in knowing which arithmetic problems are most difficult, in assessing accurately the degree of learning at any particular time, and in identifying the strategies that children use during the course of acquisition. Several studies of the development of arithmetic skills have employed the verification task to estimate differences in problem difficulty at different grade levels and to identify when children stop relying on procedural strategies such as counting and switch to fact retrieval predominantly (e.g., Ashcraft & Fierman, 1982; Ashcraft et al., 1984; Hamann & Ashcraft, 1985). The present results raise the possibility that verification performance may overestimate a child's ability to generate a solution. It appears that presenting the correct answer primes that response in memory and tends to reduce

the probability of a retrieval error and to speed RT, especially for more difficult problems. Thus, the level of retrieval skill demonstrated by a child tested with a verification task may imply a better facility with the number facts than a production task would indicate.

Furthermore, because of priming, it is possible that the strategy a child chooses for solving a given problem will be different in production and verification tasks. Siegler (1986a; Siegler & Shrager, 1984) has shown that the probability that a procedural strategy (e.g., counting) will be used for a particular problem is directly related to the speed and accuracy of retrieval on that problem. In Siegler's model, a retrieval process evaluates the set of response candidates activated by a problem (the distribution of associations), and an answer is retrieved and stated only if it exceeds a specific criterion of strength (i.e., level of activation). If none of the evaluated candidate answers exceeds the criterion threshold, the child engages a procedural backup strategy to solve the problem. Thus, the use of backup strategies depends upon the current state of the distribution of answers associated with a problem.

As explained above, the effects of priming the correct answer can be understood in terms of a temporary increase in the activation of the correct answer relative to the activation levels of false answers. In terms of Siegler's model, priming temporarily changes the "shape" of the distribution of associations, and in children it is this distribution that appears to determine whether a retrieval strategy or some other procedural backup strategy is selected. When a child is tested under true verification, the priming of the correct answer would sometimes drive the level of activation of this response above the retrieval threshold, which would result in the use of the retrieval strategy. In the absence of such priming, however, such as in a production task, the probability of the use of a procedural strategy would be higher. Thus, not only can the verification task change the apparent difficulty of problems, it conceivably has a direct impact on the way children will solve simple arithmetic problems. Consequently, studies of the development of arithmetic skill using verification run the risk of detecting transitions in strategy choice before learning is mature enough to support such transitions in the production task.

Given the evidence in Experiment 1 that, relative to the production task, presenting the correct answer in verification affected performance differently for easy and difficult multiplication problems, the question of why Ashcraft et al. (1984) found no such interaction in mental addition beyond the first grade arises. There are two main possibilities. Either multiplication and addition differ in such a way that addition is less sensitive to priming effects, or the interaction of task and difficulty exist in Ashcraft et al.'s data, but their analyses failed to detect it. Ashcraft et al. analyzed the easy-difficult factor using problem size and Wheeler's (1939) difficulty index. As explained earlier, these analyses may not maximize the chances of finding significantly different patterns of

problem difficulty for the two tasks. Moreover, Ashcraft et al. (1984) did not include the problem size (easy-difficult) factor in their analysis of errors in the two tasks. It is possible that task and difficulty did interact on accuracy measures.

It is also possible, however, that multiplication and addition interact differently with the factors present in the verification task. One possibility is that the priming effect of presenting the correct answer in verification may be less effective in addition because many of the answers are primed repeatedly by retrieval in a block of trials. For example, the sum 11 would be retrieved via $2 + 9$, $3 + 8$, $4 + 7$, $5 + 6$ (or via their reverse forms) in the course of a trial block. In contrast, in the range of multiplication problems from 2×2 to 9×9 (those tested in Experiments 1 and 2), no product is correct for more than two problems, and 26 of the 31 products in this range are uniquely correct for one problem (if, e.g., $2 \times 7 = 14$ and $7 \times 2 = 14$ are treated as the same problem). In a block of addition trials, priming due to repeated retrievals might prime certain answers (those that are correct answers to several problems) to a point at which priming by the visual route would have little additional effect.

A general implication of retrieval-priming effects (Campbell, 1987) is that neither production nor verification reveal "true" problem difficulty. Performance on any particular problem will always depend on contiguous or antecedent events that affect the activation pattern encountered in the retrieval structure when a problem is encoded. Both production and verification appear to induce priming effects within or across trials that affect performance. Thus, the conclusion to be drawn from this paper is not that verification is the villain and production the champion of cognitive arithmetic research, but rather that the two tasks are not identical with respect to factors that affect retrieval. Most generally, the present results indicate that it cannot be assumed confidently that a pattern of RTs or errors obtained in true verification will also be observed in a production task.

REFERENCES

- ANDERSON, J. R. (1981). Interference: The relationship between response latency and response accuracy. *Journal of Experimental Psychology: Human Learning & Memory*, *7*, 326-343.
- ANDERSON, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning & Verbal Behavior*, *22*, 261-295.
- ASHCRAFT, M. H. (1982). The development of mental arithmetic: A chronometric approach. *Developmental Review*, *2*, 213-236.
- ASHCRAFT, M. H. (1986). Children's knowledge of simple arithmetic: A developmental model and simulation. In J. Bisanz, C. J. Brainerd, & R. Kail (Eds.), *Formal models in developmental psychology*. New York: Springer-Verlag.
- ASHCRAFT, M. H., & FIERMAN, B. A. (1982). Mental addition in third, fourth, and sixth graders. *Journal of Experimental Child Psychology*, *33*, 216-234.
- ASHCRAFT, M. H., FIERMAN, B. A., & BARTOLOTTA, R. (1984). The production and verification tasks in mental addition: An empirical comparison. *Developmental Review*, *4*, 157-170.
- ASHCRAFT, M. H., KOSHMIDER, J. W., ROEMER, J. M., & FAUST, M.

- (1985, November). *Automaticity and practice in mental arithmetic*. Paper presented at the meeting of The Psychonomic Society, Boston.
- ASHCRAFT, M. H., & STAZYK, E. H. (1981). Mental addition: A test of three verification models. *Memory & Cognition*, **9**, 185-197.
- BAROODY, A. J. (1984). A re-examination of mental arithmetic models and data: A reply to Ashcraft. *Developmental Review*, **4**, 148-156.
- BAROODY, A. J. (1985). Mastery of the basic number combinations: Internalization of relationships or facts? *Journal for Research in Mathematics Education*, **16**(2), 83-98.
- CAMPBELL, J. I. D. (1985). *Associative interference in mental computation*. Unpublished doctoral dissertation, University of Waterloo, Waterloo, Ontario.
- CAMPBELL, J. I. D. (1987). Network interference and mental multiplication. *Journal of Experimental Psychology: Learning, Memory & Cognition*, **13**, 109-123.
- CAMPBELL, J. I. D., & GRAHAM, D. J. (1985). Mental multiplication skill: Structure, process and acquisition. *Canadian Journal of Psychology*, **39**, 338-366.
- CORBETT, A. T., & WICKELGREN, W. A. (1978). Semantic memory retrieval: Analysis of speed accuracy tradeoff functions. *Quarterly Journal of Experimental Psychology*, **30**, 1-15.
- ERIKSEN, C. W., & O'HARA, W. P. (1982). Are nominal same-different matches slower due to differences in level of processing or to response competition? *Perception & Psychophysics*, **32**, 335-344.
- GILLUND, G., & SHRIFFRIN, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, **91**, 1-68.
- GONZALEZ, E. G., & KOLERS, P. A. (1982). Mental manipulation of arithmetic symbols. *Journal of Experimental Psychology: Learning, Memory & Cognition*, **8**, 308-319.
- GRAHAM, D. J. (1987). An associative retrieval model of arithmetic memory: How children learn to multiply. In J. Sloboda & D. Rogers (Eds.), *Cognitive processes in mathematics* (pp. 123-141). Oxford, England: Oxford University Press.
- HAMANN, M. S., & ASHCRAFT, M. A. (1985). Simple and complex mental addition across development. *Journal of Experimental Child Psychology*, **40**, 49-72.
- KRUEGER, L. E. (1986). Why $2 \times 2 = 5$ looks so wrong: On the odd-even rule in product verification. *Memory & Cognition*, **14**, 141-149.
- KRUEGER, L. E., & HALLFORD, E. W. (1984). Why $2 + 2 = 5$ looks so wrong: On the odd-even rule in sum verification. *Memory & Cognition*, **12**, 171-180.
- LEFEVRE, J. A., BISANZ, J., & MRKONJIC, L. (1986, June). *Obligatory activation of arithmetic facts*. Paper presented at the meeting of the Canadian Psychological Association, Toronto.
- MATTHEWS, S. (1985). *Associative interference in adult mental addition*. Unpublished BA honors thesis, University of Waterloo, Waterloo, Ontario.
- MILLER, K., PERLMUTTER, M., & KEATING, D. (1984). Cognitive arithmetic: Comparison of operations. *Journal of Experimental Psychology: Learning, Memory & Cognition*, **10**, 46-60.
- NOREM, G. M., & KNIGHT, F. B. (1930). The learning of the 100 multiplication combinations. In G. M. Whipple (Ed.), *National Society for the Study of Education: Report on the Society's committee on arithmetic* (Vol. 15, pp. 551-567). Bloomington: Public School Publishing.
- PARKMAN, J. M. (1972). Temporal aspects of simple multiplication and comparison. *Journal of Experimental Psychology*, **95**, 437-444.
- PARKMAN, J. M., & GROEN, G. J. (1971). Temporal aspects of simple addition and comparison. *Journal of Experimental Psychology*, **89**, 335-342.
- PROCTOR, R. W. (1981). A unified theory of matching-task phenomena. *Psychological Review*, **88**, 291-326.
- REDER, L. M. (1982). Plausibility judgment versus fact retrieval: Alternative strategies for sentence verification. *Psychological Review*, **89**, 250-280.
- RUNDUS, D. (1973). Negative effects of using list items as recall cues. *Journal of Verbal Learning & Verbal Behavior*, **12**, 42-50.
- SIEGLER, R. S. (1986a). *Strategy choice procedures and the development of multiplication skill*. Unpublished manuscript, Carnegie-Mellon University, Department of Psychology, Pittsburgh.
- SIEGLER, R. S. (1986b). Unities in strategy choices across domains. In M. Perlmutter (Ed.), *Perspectives on intellectual development: Minnesota Symposium on Child Development* (Vol. 19, pp. 1-48). Hillsdale, NJ: Erlbaum.
- 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. 229-294). Hillsdale, NJ: Erlbaum.
- STAZYK, E. H., ASHCRAFT, M. H., & HAMANN, M. S. (1982). A network approach to simple multiplication. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **8**, 320-335.
- WHEELER, L. R. (1939). A comparative study of the difficulty of the 100 addition combinations. *Journal of Genetic Psychology*, **54**, 295-312.
- WICKELGREN, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, **41**, 67-85.
- WINKELMAN, J. H., & SCHMIDT, J. (1974). Associative confusions in mental arithmetic. *Journal of Experimental Psychology*, **102**, 734-736.
- ZBRODOFF, N. J., & LOGAN, G. D. (1986). On the autonomy of mental processes: A case study of arithmetic. *Journal of Experimental Psychology: General*, **115**, 118-131.

NOTES

1. At the time Experiment 1 was designed, no normative data on adults' specific multiplication errors had been published. Two experiments conducted prior to Experiment 1 provided the error-frequency data used to identify common false answers for each problem. In one pilot study, each of 10 subjects was tested on each problem 12 times, and in the other, 10 subjects were tested on each problem 10 times. Together, the experiments provided about 400 errors. Recently, Campbell and Graham (1985, Appendix A) presented a table of frequencies of specific multiplication errors based on a sample of 60 adults. As is shown below, the particular false relationships identified from the pilot studies are substantially confirmed by a comparison with the error matrix presented by Campbell and Graham.

2. There were two exceptions. The problem 2×3 was assigned 7 as its table-unrelated answer, and 2×2 was assigned 5 as its unrelated product. In each case, the answer is a table-unrelated product (i.e., from the times-one table), but not an answer to one of the problems tested.

3. Krueger (1986) presented evidence that subjects can efficiently use an odd-even rule on false-verification trials in multiplication as an initial plausibility check of the presented answer (i.e., if either multiplier is even, the answer must be even). Although the possible effects of the odd-even status of answers were not anticipated when the present experiment was designed, fortuitously, the two sets of false answers within each of the low-error and high-error problem sets are well balanced in terms of the number of equations satisfying and violating the odd-even rule (see Appendix A). Moreover, Krueger proposed that use of the odd-even rule is a strategic decision on the part of subjects and that the ready accessibility of correct answers for simple arithmetic problems often preempts plausibility-judgment strategies (see also Reder, 1982). Instead, subjects generally adopt a retrieval strategy. This seems especially likely in the present experiment. Less than 19% of trials (including true and false types) involved an answer in violation of the odd-even rule. Thus, an odd-even check would prove uninformative on over 80% of the trials.

4. The low-error versus high-error factor represents a classification of problems into a set that subjects in the pilot studies tended to find "easy" and a set that their performance indicated was more "difficult." It might be possible to define two new sets using the normative multiplication data in Campbell and Graham (1985) in order to maximize differences in normative difficulty between the sets. As pointed out earlier, the Campbell and Graham data generally confirm the present breakdown, and preserving the initial assignment permits some informal comparisons with the false-answer conditions for the two sets. Note, however, that in the false conditions the low-error versus high-error distinction is confounded with pairings of different types of false answers and corresponds to two different conditions for testing the confusion-product effect. This means that a formal analysis comparing true and false verification for the low- and high-error sets is not meaningful. Generally, however, true and false trials yielded similar error rates

overall, and, as is usually found (e.g., Ashcraft & Stazyk, 1981; Parkman, 1972), false verification trials yielded longer RTs than did true verification.

5. The correct product is typically found to be the problem-size variable correlating most highly with problem RTs in multiplication when times-one problems are included and tie problems (e.g., 4×4 , 7×7 , etc.) are excluded from the analysis (e.g., Miller et al., 1984; Stazyk et al., 1982). Problems involving 0 or 1 were not tested in the present experiment because it is likely that these are solved by rule-based procedures rather than by retrieval of answers (cf. Baroody, 1985). The finding that max was the problem-size variable yielding the highest correlation in the present study is not anomalous, but simply reflects the particular set of problems (i.e., 2×2 through 9×9) that went into the analyses. Campbell (1985) found that over this set of problems, max was correlated highest with problem RT for 45 of 60 subjects and that the correct product was correlated highest for only 3 of the subjects. In the present context, no specific theoretical significance should be attached to max beyond the fact that difficulty tends to increase with problem size.

6. There are some discrepancies between Experiments 1 and 2 in the assignments of false answers. For example, in Experiment 1, 49 was identified as a LF error for 7×9 but appears as the related HF error answer for 7×9 in Experiment 2. This occurred because different sets of error data were used to identify false answers for the two experiments. As shown before, the assignments used in Experiment 1 were largely confirmed by a comparison with Campbell and Graham's (1985) norms, and the pattern of production errors in Experiment 1 clearly demonstrated that the assignments accomplished their general purpose. Nonetheless, the pilot data used for designing Experiment 1 was based on one third as many subjects as contributed to the Campbell and Graham error matrix, and the latter is undoubtedly the more reliable data base.

7. Mutual exclusion was not used for the initial ANOVA because applying the exclusion rule across four conditions leads to a serious depletion of data. Means in each of the four conditions were computed for each subject using all the correct RTs (excluding outliers) available within each condition.

8. The four cells defined by the neutral versus correct and easy versus difficult factors were not explicitly balanced for practice effects. To check that randomization produced an acceptable balancing of the four conditions, the mean testing position (i.e., the mean block across subjects in which a problem was tested) was computed for each problem for the correct and neutral prime conditions. If, across subjects, each problem was tested equally often in each block in each condition, the mean testing position for each item in each condition would be 2.5 (i.e., the sum of Block Positions 1 through 4 divided by the number of blocks). On the basis of the mean positions for each problem, an average testing position and standard deviation were computed for each of the four cells. The average of the four cell means was 2.52 ($SD = .05$). The average of the standard deviations of item-position means in each cell was .31 ($SD = .05$).

9. The means and standard deviations (including the data from 7×9) for the neutral, unrelated, and related priming conditions, respectively, were: 835 (100), 857 (115), and 891 (105). The 34-msec difference between related and unrelated was reliable [$t(35) = 2.52, p < .02$]; however, the 22-msec difference between neutral and unrelated was not significant [$t(35) = 1.68, p > .05$].

10. The ANOVA testing the related versus unrelated \times easy versus difficult interaction is not reported here because of low reliability of the RT means in the difficult cells (after mutual exclusion, the rate of missing data was 38% overall, and for several subjects the rate approached 70%). It should be pointed out that the ANOVA indicated no differences between related and unrelated primes for the difficult problem set. Although regression using problem means is probably more reliable and revealed no significant interaction, this outcome should be interpreted cautiously.

APPENDIX A The Low-Error and High-Error Problems and Their Verification Answers in Experiment 1

20 Low-Error Problems				16 High-Error Problems			
PRB	True	Rel	Unr	PRB	True	LF-Err	HF-Err
2×2	4	6	5	2×6	12	10	18
2×3	6	9	7	2×8	16	24	18
2×4	8	6	9	3×8	24	16	18
2×5	10	8	9	3×9	27	21	18
2×7	14	12	15	4×6	24	28	20
2×9	18	27	25	4×7	28	32	24
3×3	9	6	8	4×8	32	28	24
3×4	12	15	14	4×9	36	27	32
3×5	15	10	16	5×8	40	32	30
3×6	18	15	14	5×9	45	36	35
3×7	21	28	25	6×7	42	30	32
4×4	16	20	10	6×8	48	36	42
4×5	20	16	21	6×9	54	48	45
5×5	25	20	21	7×8	56	40	42
5×6	30	36	28	7×9	63	49	56
5×7	35	42	27	8×9	72	54	56
6×6	36	42	32				
7×7	49	56	54				
8×8	64	56	54				
9×9	81	72	64				

Note—PRB = problem; True = the correct product; Rel = a false product from the same times-table as the problem (table related); Unr = a false product not from the same times-table (table unrelated); LF-Err = LF table-related error; and HF-Err = HF error.

APPENDIX B Easy and Difficult Problem Sets and the False Answers Assigned to Each Problem for Experiment 2

Easy-Problem Set				Difficult-Problem Set			
Problem	Cor	Rel	Unr	Problem	Cor	Rel	Unr
2×2	4	6	9	2×8	16	18	15
2×3	6	8	7	3×6	18	12	14
2×4	8	6	9	3×8	24	18	14
2×5	10	8	9	3×9	27	18	16
2×6	12	18	15	4×6	24	28	27
2×7	14	21	20	4×7	28	24	25
2×9	18	16	15	4×8	32	24	21
3×3	9	6	8	4×9	36	32	35
3×4	12	24	25	5×7	35	45	48
3×5	15	18	14	5×8	40	45	42
3×7	21	27	25	5×9	45	40	48
4×4	16	8	9	6×7	42	49	40
4×5	20	10	18	6×8	48	42	49
5×5	25	35	32	6×9	54	36	32
5×6	30	35	32	7×8	56	48	45
6×6	36	42	45	7×9	63	49	48
7×7	49	42	45	8×8	64	48	42
9×9	81	36	30	8×9	72	48	49

Note—Cor = correct answer; Rel = false table related (HF error); Unr = false table unrelated (LF error).