

Why is $9 + 7$ harder than $2 + 3$? Strength and interference as explanations of the problem-size effect

N. JANE ZBRODOFF

University of Illinois at Urbana-Champaign, Urbana, Illinois

In four experiments, the problem-size effect was investigated, using an alphabet-arithmetic task in which subjects verified such problems as $A + 2 = C$. Problem size was manipulated by varying the magnitude of the digit addend (e.g., $A + 2$, $A + 3$, and $A + 4$). The frequency and similarity of problems was also manipulated to determine the contribution of strength and interference, respectively. Experiment 1 manipulated frequency at low levels of practice and found that strength could account for the problem-size effect. Experiment 2 manipulated frequency at higher levels of practice, and found that strength alone could not account for the problem-size effect at asymptote. Experiment 3 manipulated frequency and similarity and found a substantial problem-size effect at asymptote, suggesting that both strength and interference contribute to the problem-size effect. Experiment 4 manipulated similarity, keeping frequency constant, and found no problem-size effect at asymptote, suggesting that interference alone is not responsible for the problem-size effect. The results are related to findings with number arithmetic.

One of the most fundamental findings in the research on simple arithmetic is that people take longer and make more errors when they solve large problems (i.e., problems with large digits) than when they solve small problems (i.e., problems with small digits). This is most commonly referred to as the *problem-size effect*. It occurs with children and with adults, with addition, subtraction, multiplication, and division problems, and with production and verification tasks. Much effort has gone into trying to explain the problem-size effect. In effect, all theories of simple arithmetic try to account for it, and how well they are able to account for it determines how good they are as theories.

There are four types of theories that may explain the problem-size effect. Two are reconstructive theories and two are reproductive theories. Reconstructive theories propose an algorithm that computes the answer; reproductive theories propose a retrieval process that retrieves previous answers. The time taken by the algorithm or the retrieval process depends on the number or the difficulty of the steps to be executed. The theories argue that larger problems require more steps or more difficult steps and therefore take longer to solve and are more prone to errors. The question is, why does the number or the difficulty of steps depend on the magnitude of the problem? The answers to this question reveal the fundamental assumptions that distinguish the theories.

Reconstructive Theories

Number of Steps

The primary reconstructive theories are counting theories. They assume that answers are computed by setting an internal counter to a specified value and then successively incrementing it a specified number of times. Five counting models have been investigated (Groen & Parkman, 1972). In the sum model, the counter is set to zero and then incremented by both digits in the problem; in the right model, the counter is set to the left digit and then incremented by the right digit; in the left model, the counter is set to the right digit and then incremented by the left digit; in the max model, the counter is set to the minimum digit and then incremented by the maximum digit; in the min model, the counter is set to the maximum digit and then incremented by the minimum digit.

All the counting models assume that the number of times an internal counter is incremented depends on the magnitude of the digit(s) used to increment the counter. As the magnitude increases, the number of increments increases, and as the number of increments increases, the longer it will take and the more opportunity there will be to make an error. However, for any particular problem, each model predicts a different number of increments and therefore a different solution time. The one that accounts for solution times best is the min model. In experiments with both children and adults and with addition and multiplication, it has been found that reaction times increase linearly with the magnitude of the minimum digit (Groen & Parkman, 1972). It has been found that even children as young as 4 use the min model to solve addition problems (Groen & Resnick, 1977).

It seems intuitive that young children solve addition problems by counting, but it is not so intuitive that adults do. Given the amount of practice adults have had in solv-

These data were reported at the 35th Annual Meeting of the Psychonomic Society in St. Louis, November 1994. I would like to thank Julie Delheimer for testing subjects and analyzing data and Mark Ashcraft, Jamie Campbell, Gordon Logan, and Michael Masson for helpful comments. Correspondence concerning this article may be addressed to N. J. Zbrodoff, Department of Psychology, University of Illinois, 603 East Daniel St., Champaign, IL 61820 (e-mail: jzbrodof@s.psych.uiuc.edu).

ing simple addition problems, it is likely that they have memorized the facts. Thus, although they may on occasion use a counting strategy (Svenson, 1985), it is more likely that they typically retrieve the facts. Children who are acquiring arithmetic skill report counting initially, but reports of counting decline and reports of remembering increase as skill increases (Siegler, 1987; Zbrodoff, 1979). Adults who are acquiring analogs of arithmetic (i.e., alphabet arithmetic; see below) report a similar transition from counting to remembering (Compton & Logan, 1991; Logan & Klapp, 1991). Moreover, there are a number of associative facilitation and interference effects that are best accounted for by retrieval theories (e.g., Campbell, 1987, Campbell & Clark, 1989; Zbrodoff & Logan, 1986, 1990).

Difficulty of Steps

No current reconstructive theory of simple arithmetic proposes an algorithm in which the difficulty of the steps is the primary determinant of the problem-size effect. Thomas (1963) proposed a model in which the time to solve simple and complex arithmetic problems depended on the sum of the logarithms of the digits and the answer. However, he did not specify the mechanism underlying the algorithm and did not explain why the difficulty of the computation depended on problem size.

Retrieval Theories

Number of Steps

Ashcraft and Battaglia (1978) and Geary, Widaman, and Little (1986) proposed theories in which retrieval was a sequential table-search process in which the number of steps depended on the size of the problem. Arithmetic facts are represented as tables in which the rows correspond to one of the digits, the columns correspond to the other digit, and the entries in the table are the sums or products of the digits. Retrieval involves entering the table at 0,0 (or 1,1), moving down the rows to find the first digit, moving across the columns to find the second digit, and retrieving the sum or product that is at the intersection of the row and column. Thus, the total number of steps required depends on the sum of the digits. Geary et al. (1986) argued that reaction time was a linear function of the number of steps. Ashcraft and Battaglia (1978) argued that the steps became more difficult for larger digits and suggested that reaction time should increase with the square of the sum. These models have received some support. In adult data, the sum or the sum squared often predicted reaction time as well as or better than the minimum digit (Ashcraft & Battaglia, 1978; Geary et al., 1986).

These models assume that verification tasks (e.g., $4 + 3 = 12$, true or false?) involve a table-search retrieval stage plus a comparison stage. First, an answer is retrieved from the table (i.e., 7), and then it is compared with the presented answer (i.e., 12). If it is the same as the presented answer, it is true, and if it differs, it is false. Table-search models predict that the nature of the presented answer should not affect the retrieval stage since retrieval is completed before comparison begins. The problem is that there are findings that suggest that presented answers do affect retrieval

(e.g., Campbell, 1987; Campbell & Clark, 1989; Winkelman & Schmidt, 1974; Zbrodoff, 1979; Zbrodoff & Logan, 1986, 1990).

Work by Winkelman and Schmidt (1974) and others (Zbrodoff, 1979; Zbrodoff & Logan, 1986, 1990) suggests that presented answers produce associative confusion effects if they are answers that are correct for another arithmetic operation. For example, problems like $4 + 3 = 12$ take longer to reject than problems like $4 + 3 = 11$. Presumably, 12 rings true when it is presented because it is the correct answer for a multiplication operation, and it takes extra time to overcome the tendency to say "true" to it. On the other hand, 11 is not correct for any operation so it does not interfere with the retrieval of 7. Work on error priming by Campbell (1987, 1990, 1991) also suggests that answers may affect retrieval. For example, answers to previous problems often appear as errors on later problems. Table-search models cannot account for either associative confusion or error-priming effects.

Table-search models also predict that problem-size effects in verification should decrease as the delay between onset of retrieval and onset of comparison increases. This is because the more time there is for retrieval, the more likely it is to have been completed before the onset of comparison. At long delays, even the most difficult retrievals should be completed before the answer is presented, so the time to make the comparison should not be affected by the size of the problem. This prediction was tested in addition and multiplication, using both experimenter-imposed and subject-imposed delays (Zbrodoff & Logan, 1990). Even at the longest subject-imposed delays, there was still a substantial problem-size effect. Once again, this suggests that the presented answer must influence the retrieval stage, contrary to what table-search models assume.

Difficulty of Steps

Associative theories of retrieval assume that arithmetic facts are represented by associations between digits and answers. When a problem is presented, the digits act as cues that retrieve the answer that is associated with the digits. There are two types of associative theories: strength and interference theories. They both assume that there is a single-step retrieval process regardless of the size of the problem but that the difficulty of retrieval increases as the size of the problem increases.

Strength theories assume that the difficulty of retrieval depends on the strength of the association between digits and answers, and interference theories assume that it depends on the amount of interference from competing associations. Some current theories explain the problem-size effect primarily in terms of strength (Ashcraft, 1992); others explain it primarily in terms of interference (Campbell & Oliphant, 1992; Siegler, 1986, 1988). The present work explores the possibility that a combination of strength and interference may account for the problem-size effect.

Strength theories. In strength theories, the difficulty of retrieval depends on the strength of association between the digits and the answer. The strength of association increases with its frequency of use, and the stronger the as-

sociation, the faster it is retrieved. If small problems are more frequent than large problems, then strength theories can explain the problem-size effect. Small problems would have stronger associations than large problems and would be retrieved faster than large problems. Several analyses of elementary-school mathematics texts have revealed that small problems do in fact appear more often than large problems (Ashcraft & Christy, *in press*; Clapp, 1924; Hamann & Ashcraft, 1986; Thorndike, 1922). Moreover, there is a general tendency for small numbers to occur more often than large numbers in many domains. For example, Benford (1938) found that small numbers occur more frequently than large numbers as the first digit in an entry in more than 20 tables of physical measurements (e.g., molecular weights of chemical compounds, surface areas, street addresses). Specifically, Benford's (1938) law states that the frequency of numbers appearing as the first digit in an entry decreases as a logarithmic function of their size. The frequency of small versus large problems may explain the problem-size effect in elementary-school children, for whom differences in frequency have been documented, but it may not explain it in adults. The distribution of frequencies may be different for adults and children. No one has measured the frequency with which problems occur in adult life, although Benford's law suggests that small problems may occur more often than large ones throughout life. Even if small problems occur more often than large ones in adult experience, overall, both have occurred so often that their associative strengths may be at asymptote and differences in their frequency may no longer make much of a difference. Differences in frequency may affect the rate at which performance approaches asymptote more than they affect the level of performance at asymptote. This is an empirical question that is explored in the present work.

Interference theories. In interference theories of arithmetic, the difficulty of retrieving an answer depends on the interference it suffers from other answers. The more interference there is, the harder it is to retrieve the correct answer. Siegler (1988; see also Siegler & Shrager, 1984) argued that there might be interference from incorrect associations formed during the acquisition of arithmetic skill that continues throughout life. People make more errors learning large problems than they do learning small problems, so large problems have more incorrect associations that interfere with retrieval than do small problems.¹ Also, large problems may not have been initially learned as well as small problems. Because large problems are typically learned later than small problems, they may have been subject to more proactive interference from previously learned problems than would small problems (Campbell & Graham, 1985). Answers to large problems may also be less discriminable from answers to other large problems than answers to small problems may be from answers to other small problems (Campbell & Oliphant, 1992). The less the discriminability, the more interference, and thus the more interference there may be with large problems than with small problems.

Interference also depends on the similarity between problems. A given problem will retrieve answers that are associated with its operands, but it may also retrieve answers to other problems that are similar to it. Problems are similar if they share some physical or numerical feature. In particular, they are similar if they share an operand, have the same operands but a different operation sign, or share a number in the 1s or 10s value of their answer. For example, the problem $2 + 3$ will retrieve 5, but it may also retrieve 4 and 6 because it shares an operand with the problems $2 + 2$ and $3 + 3$. It may also retrieve 6 because $2 \times 3 = 6$. The problem 7×7 may retrieve 45 as well as 49 because both answers have a 4 in the 10s column. Presumably, the greater the number of similar problems, the greater the number of competing associations that may interfere with the retrieval of a correct answer. While similarity-based interference can explain some of the common types of errors that occur (e.g., table-related errors, cross-operation errors), it cannot explain the problem-size effect. This is because the potential for similarity-based interference is equal for large and small problems across the entire set of addition and multiplication facts. The problem $2 + 3$ has just as many problems that share an operand with it as the problem $9 + 7$ has, for example.

In the present experiments the possibility that a combination of similarity-based interference and strength might explain the problem-size effect was investigated. The idea is that given a potential for interference, strong associations may have a greater potential to interfere with the retrieval of weak associations than vice versa. If small problems are simply more frequent than large problems, then they should have a greater strength and thus a greater potential to interfere than large problems. For example, if $7 + 2 = 9$ is stronger than $7 + 9 = 16$, then $7 + 2 = 9$ is more likely to be retrieved when $7 + 9$ is presented than $7 + 9 = 16$ is likely to be retrieved when $7 + 2$ is presented. And if $7 + 3 = 10$, $7 + 4 = 11$, $7 + 5 = 12$, $7 + 6 = 13$, $7 + 7 = 14$, and $7 + 8 = 15$ are all stronger than $7 + 9 = 16$, they all may be more likely to be retrieved when $7 + 9$ is presented than when $7 + 2$ is presented. Hence, the retrieval of answers to large problems may involve more competition than the retrieval of answers to small problems. Moreover, if the resolution of interference depends on the difference in strength between the competing associations, the interference will be harder to resolve for large problems than for small problems. With large problems, a relatively weak correct answer competes with relatively strong incorrect answers, whereas with small problems, a relatively strong correct answer competes with relatively weak incorrect answers.

A combination of interference and strength may explain the problem-size effect in children, where there is evidence that large and small problems differ in strength. However, it may not account for the effect in adults. If interference depends on a difference in strength between competing associations, then the extensive practice adults have with arithmetic may eliminate the difference. Large problems may become as strong as small problems, so they may be able to overcome interference from compet-

ing traces just as easily as small problems. Whether there are different interference effects for large and small problems as performance reaches asymptote is tested in the present work.

THE EXPERIMENTS

In the present experiments, frequency and similarity were manipulated independently to distinguish between strength and interference theories of the problem-size effect. In Experiments 1 and 2, the size and frequency of problems was varied systematically and the similarity between problems was kept consistently low. In Experiment 3, the size and frequency of problems was varied and the similarity between problems was kept consistently high. In Experiment 4, the size of problems was varied, the frequency of problems was kept constant, and the similarity between problems was kept consistently high.

The present experiments employed an analog of arithmetic—*alphabet arithmetic*—that was devised by Logan and his colleagues (Compton & Logan, 1991; Logan & Klapp, 1991) to study automaticity in adults. The alphabet-arithmetic task requires subjects to verify such problems as $B + 2 = E$ or $M + 4 = Q$. Subjects report counting through the alphabet in order to solve the problem. In the first example, they start at B and count two steps down the alphabet to D. Then they compare D with the presented answer, E, and respond “false.” In the second example, they start at M, count four steps to Q, compare the “counted” Q with the presented answer, Q, and respond “true.” Consistent with these reports, verification times increased linearly with the magnitude of the digit addend, which determines the number of counting steps. The slope of the function was typically 400–500 msec per count. After extended practice with the problems, subjects report relying on memory rather than counting to solve the problems, and consistent with these reports, the slope of the function relating verification time to the digit addend is close to zero.

The present experiments used alphabet arithmetic with adults to mimic the acquisition of numerical arithmetic in children. Adults confront the alphabet-arithmetic task with skills that are analogous to those that children have when they confront numerical addition for the first time. Adults know the alphabet sequence just as children know the number sequence and both know how to count. Thus, adults’ initial performance on the alphabet-arithmetic task may reflect processes that are similar to children’s initial performance. Moreover, adults’ transition from the counting algorithm to memory retrieval may reflect processes similar to those underlying the transition in children. However, the transition can be controlled much better with alphabet arithmetic in adults than with regular arithmetic in children. The transition with alphabet arithmetic can be produced in a single session, if necessary (see, e.g., Logan & Klapp, 1991, Experiment 2), whereas the transition with regular arithmetic may span several years (Siegler, 1986; Zbrodoff, 1979). Alphabet arithmetic is especially useful in the present context because the frequency with

which problems are presented can be manipulated independently of the similarity between problems. It is very difficult, for practical and ethical reasons, to manipulate frequency and similarity independently in number arithmetic with children.

Experiment 1

Experiment 1 tested a prediction of strength theories of the problem-size effect. According to strength theories, the problem-size effect occurs because small problems are practiced more frequently and are therefore stronger than large problems. Thus, it should be possible to reverse the problem-size effect by presenting large problems more frequently than small problems. The size of problems was manipulated by varying the magnitude of the digit addend. Digit addends were 2, 3, or 4. One group of subjects, the *standard* group, received more practice with small problems than with large problems, mimicking the “standard” frequency imbalance in children’s arithmetic texts. Specifically, problems with addend 2 occurred 48 times in a block of trials, problems with addend 3 occurred 32 times, and problems with addend 4 occurred 16 times. This group should show a problem-size effect throughout practice. The *reverse* group received more practice with large problems than with small problems. Specifically, problems with addend 2 occurred 16 times in a block, problems with addend 3 occurred 32 times, and problems with addend 4 occurred 48 times. This group should show a reversed problem-size effect after the transition from counting to retrieval. The third group, the *control* group, received large and small problems equally frequently (i.e., problems with digit addends of 2, 3, and 4 each occurred 32 times per block). This group should show no problem-size effect after the transition to retrieval.

Method

Subjects. Fifty-four undergraduate students at the University of Illinois participated in Experiment 1 in order to earn credit in their introductory psychology course.

Apparatus and Stimuli. Alphabet-arithmetic problems were constructed by pairing the 18 letters A through R with the addends, 2, 3, and 4, producing 54 combinations. Nine groups of problems were constructed by pairing three sets of consecutive letters (i.e., A–F, G–L, M–R) with three digit addends (i.e., 2, 3, 4). Within a group, each letter was paired with only one digit but each digit was paired with two letters. An example of a group is: A + 2, B + 3, C + 4, D + 2, E + 3, F + 4. Across groups, each digit was paired with each letter equally often. Each problem in a group was paired once with the true answer and once with a false answer, which was the true answer + 1.

Problems were presented as white characters on a dark background in the middle of an Amdek Model 722 color monitor controlled by IBM PC/XT or AT computers. Before each trial, there was a 500-msec warning interval in which two lines of dashes were presented, one line above and one line below the location in which the problem would appear. Then the problem appeared and remained on the screen until a response key (i.e., either the “z” key or the “/” key) was pressed. After the response, the screen went blank for a 2,000-msec intertrial interval, after which the warning signal for the next trial was presented. The computers measured time in milliseconds from the onset of the problem to the onset of a response and recorded which response was made.

Procedure. Each subject was presented with three blocks of problems. Each block contained 96 true and 96 false problems, for a total of 576 trials in a session. In the standard group, in each block, each of the two problems with the addend 2 (i.e., small problems) was presented 24 times, each of the two problems with the addend 3 (i.e., medium problems) was presented 16 times, and each of the two problems with the addend 4 (i.e., large problems) was presented 8 times. Subjects in the reverse group received each large problem 24 times, each medium problem 16 times, and each small problems 8 times in a block. Subjects in the control group received all 6 problems 16 times in a block. There were 18 subjects in each group. Two subjects in each group received the same problem sets; one pressed the “/” key and the other pressed the “z” key to indicate “true.”

The subjects were told that they would be performing a task that was like regular arithmetic but also involved letters. They were given $S + 3 = W$ as an example (it did not appear in any of the experimental problem sets) and were told that they could determine whether problems were true or false by starting at the letter given in a problem and counting through the alphabet the number of letters specified by the digit in the problem. The counting procedure was demonstrated on the example problem. They were told which keys to press for “true” and “false” responses and to respond as quickly as they could without making too many mistakes. The subjects were allowed brief breaks every 72 trials.

Results and Discussion

Mean reaction time and accuracy were computed for each combination of digit addend, truth value, and block for each subject. The means across subjects are presented in Table 1. The overall average reaction time was 1,672 msec and was not affected by group [$F(2,51) < 1$, $MS_e = 4,514,475.17$]. Accuracy was high, averaging 95.1%, and correlated negatively with reaction time [$r = -.548$, $F(1,52) = 22.35$, $p < .01$].

The key results have to do with the problem-size effect. The problem-size effect was estimated by fitting straight lines to the functions relating reaction time to the digit addend. The slope of the best-fitting straight line was used as an index of the problem-size effect. The mean slopes in each group are plotted as a function of block in Figure 1.

The control group (in which small and large problems were equally frequent) replicated standard effects in the alphabet-arithmetic literature. The problem-size effect was substantial in the first block but diminished to zero by the third block, indicating the transition from counting to retrieval observed previously (cf. Compton & Logan, 1991; Logan & Klapp, 1991). The standard group (in which small problems were more frequent than large problems) showed a stronger problem-size effect in every block. In the first block, the slopes were nearly twice as large as slopes observed in past research when subjects reported counting (Compton & Logan, 1991; Logan & Klapp, 1991, Experiment 1). Most likely, subjects there had memorized more of the frequently presented small problems than they had the infrequently presented large problems, so small problems were driven by rapid memory retrieval while large problems were driven by the slow counting process, creating a large problem-size effect. However, the standard group showed a large problem-size effect even in the third block when all performance was probably based primarily on retrieval. This is a demonstration of the strength-theory account of the problem-size effect: Small problems occur more frequently and are stronger than large problems and are therefore responded to faster than large problems.

The reverse group (in which large problems were more frequent than small problems) showed a negligible problem-size effect in the first block which became negative in Blocks 2 and 3. The reversal is important because it demonstrates that frequency, not the magnitude of the arguments, is responsible for the problem-size effect. The reversal is strong evidence for strength theories.

These conclusions were supported by analyses of variance (ANOVAs) on the mean reaction times and on the slopes. In the 3 (group: control vs. standard vs. reverse) \times 3 (block) \times 2 (truth) \times 3 (addend) ANOVA on the reaction times, the main effects of block [$F(2,102) = 225.98$, $MS_e = 736,926.42$, $p < .01$], truth [$F(1,51) = 72.88$, $MS_e =$

Table 1
Mean Reaction Times (RT, in Milliseconds) and Percentages of Correct Responses (% Corr) for Each Group in Experiment 1 as a Function of Block, Addend, and Truth

Block	Addend											
	True						False					
	2		3		4		2		3		4	
RT	% Corr	RT	% Corr	RT	% Corr	RT	% Corr	RT	% Corr	RT	% Corr	
Control Group												
1	1,728	95	2,293	93	2,727	93	1,952	96	2,629	94	2,903	92
2	1,136	96	1,345	95	1,317	97	1,281	97	1,557	95	1,531	93
3	1,077	95	1,143	96	1,131	99	1,194	96	1,281	95	1,203	98
Standard Group												
1	1,685	97	2,478	96	3,535	91	1,996	96	2,818	95	3,569	94
2	981	98	1,314	97	1,899	94	1,139	98	1,594	94	1,933	93
3	843	98	992	96	1,439	93	981	98	1,169	96	1,513	93
Reverse Group												
1	2,117	95	2,462	96	2,317	95	2,391	90	2,589	92	2,557	94
2	1,355	95	1,394	93	1,017	96	1,591	92	1,534	97	1,184	97
3	1,194	95	1,070	92	798	97	1,284	92	1,174	98	947	97

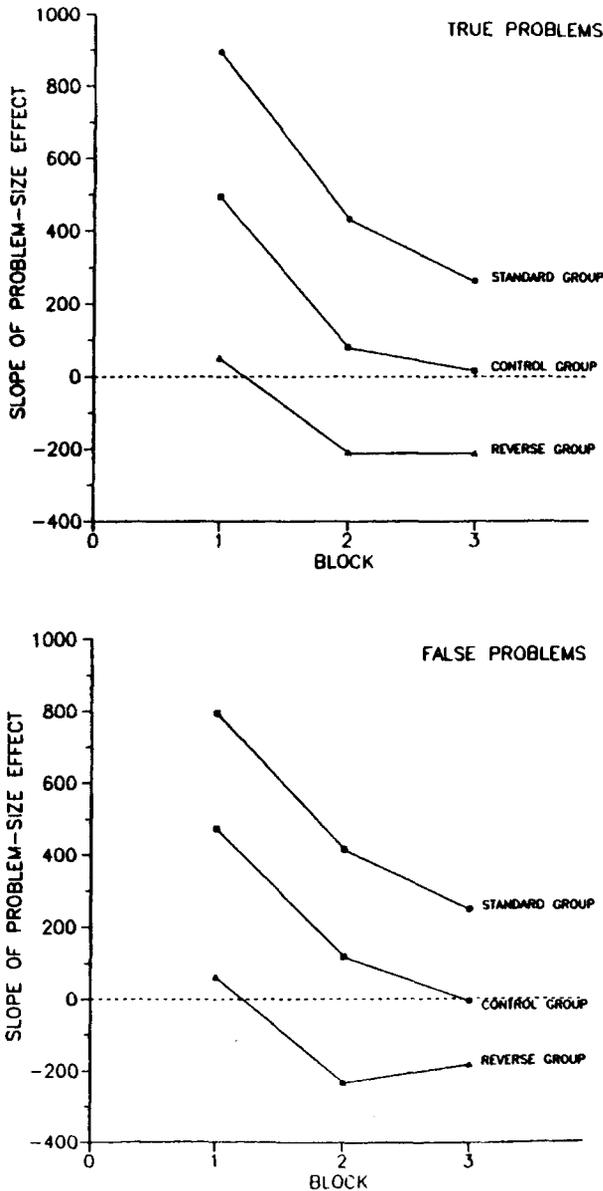


Figure 1. Mean slope of the linear function relating reaction time to digit addend as a function of practice block for true problems (top panel) and false problems (bottom panel) in Experiment 1. (Group is the parameter: control vs. standard vs. reverse.)

101,027.84, $p < .01$], and addend [$F(2,102) = 31.30$, $MS_e = 474,703.12$, $p < .01$] were significant. The interactions between block and addend [$F(4,204) = 45.04$, $MS_e = 133,981.25$, $p < .01$] and group and addend [$F(4,102) = 22.46$, $MS_e = 474,703.12$, $p < .01$] were significant. In the 3 (group: control vs. standard vs. reverse) \times 3 (block) \times 2 (truth) ANOVA on the slopes, there were significant main effects of group [$F(2,51) = 34.72$, $MS_e = 307,387.50$, $p < .01$] and block [$F(2,102) = 60.05$, $MS_e = 99,344.66$, $p < .01$] and a significant interaction between group and block [$F(4,102) = 2.66$, $MS_e = 99,344.66$, $p < .05$].

Experiment 2

Experiment 1 showed that strength theories could account for the problem-size effect at low levels of practice. Experiment 2 was designed to determine whether they could also account for the problem-size effect at high levels of practice when performance becomes asymptotic. Subjects were tested for five sessions in the standard condition of Experiment 1. That is, they received addend-2 problems 48 times per block, addend-3 problems 32 times per block, and addend-4 problems 16 times per block. They had five times as much practice as had subjects in Experiment 1. The main question was whether the problem-size effect would be significantly greater than zero when performance was at asymptote. If strength theories are to account for the problem-size effect in adults, who are likely to be performing at asymptote, then they must predict a greater-than-zero problem-size effect here. Alternatively, finding a problem-size effect of zero would suggest that some other principle, perhaps interference, was responsible for the problem-size effect in adults.

Method

The method was the same as in the standard group in Experiment 1 except that subjects performed the task for five sessions, one session a day. The subjects were 18 undergraduate students from the University of Illinois who did not participate in Experiment 1. They were paid \$20 for their participation. Full instructions were given on the first session and reviewed briefly at the beginning of each subsequent session.

Results and Discussion

Mean reaction time and accuracy were computed for each combination of digit addend, truth value, and block (there were three blocks of trials on each of five sessions) for each subject. The means across subjects are presented in Table 2. Accuracy was high, averaging 94.8%, and relatively stable over practice, whereas reaction time declined substantially from an average of 2,781 msec in Block 1 to an average of 635 msec in Block 15. Consequently, the correlation between reaction time and accuracy was not significant [$r = -.057$, $F(1,88) < 1$]. The average reaction time was 906 msec.

As in Experiment 1, the problem-size effect was estimated as the slope of a straight line fitted to the function relating reaction time to digit addend. The mean slopes in each group are plotted as a function of block in Figure 2. The slopes from Experiments 3 and 4 are plotted there as well for comparison. The problem-size effect was large in the first block of trials but diminished rapidly. It was less than 100 msec by the fifth block and less than 50 msec by the sixth block. In the final session (Blocks 13–15), it was close to zero, averaging 25 msec. These results suggest that differences in frequency do not lead to differences in asymptotic performance. This suggests that strength alone cannot account for the problem-size effect in numerical arithmetic with adults, who are presumably performing at asymptote.

These conclusions were supported in a 15 (block) \times 2 (truth) \times 3 (addend) ANOVA on the mean reaction times

Table 2
Mean Reaction Times (RT, in Milliseconds) and Percentages of Correct Responses (% Corr)
in Experiment 2 as a Function of Block, Addend, and Truth

Block	Addend											
	True						False					
	2		3		4		2		3		4	
RT	% Corr	RT	% Corr	RT	% Corr	RT	% Corr	RT	% Corr	RT	% Corr	
1	1,728	96	2,470	94	3,160	92	1,908	96	2,805	95	3,530	95
2	998	97	1,406	97	1,748	91	1,111	98	1,717	97	1,890	94
3	821	97	1,091	98	1,227	93	913	98	1,235	94	1,245	94
4	757	97	898	97	984	92	834	98	1,018	94	1,042	90
5	692	97	767	96	786	92	753	97	849	94	861	90
6	691	97	752	96	789	94	736	98	821	95	809	89
7	655	97	705	96	723	94	690	99	756	96	764	87
8	651	97	670	97	694	94	674	98	746	95	748	90
9	653	96	664	97	686	94	661	99	710	97	752	91
10	609	96	662	98	661	92	616	98	670	97	731	87
11	610	96	649	95	676	92	611	98	672	95	685	89
12	632	96	642	96	677	92	641	99	676	96	673	90
13	595	96	608	96	614	93	580	98	642	96	654	86
14	603	97	627	95	646	93	609	98	666	95	674	86
15	615	95	639	97	633	93	617	98	653	95	655	86

and a 15 (block) \times 2 (truth) ANOVA on the slopes. In the reaction time ANOVA, there were significant main effects for block [$F(14,238) = 65.48$, $MS_e = 449,106.98$, $p < .01$], truth [$F(1,17) = 51.25$, $MS_e = 36,052.24$, $p < .01$], and addend [$F(2,34) = 42.49$, $MS_e = 185,728.92$, $p < .01$] and significant interactions between block and truth [$F(14,238) = 10.44$, $MS_e = 15,473.97$, $p < .01$] and block and addend [$F(28,476) = 25.93$, $MS_e = 57,574.07$, $p < .01$]. In the slope ANOVA, only the main effect of block was significant [$F(14,238) = 39.43$, $MS_e = 39,425.66$, $p < .01$]. Fisher's least significant difference (LSD), calculated from the error term for the main effect of block (39,425.66), showed that the average slope per block was not significantly different from zero (LSD = 92, $p < .05$) from the fifth block on.

Experiment 3

Experiment 2 showed that strength alone cannot account for the problem-size effect at asymptote. Experiment 3 asked whether strength and interference together would be sufficient. In Experiments 1 and 2, subjects learned fairly dissimilar problems. They learned one fact about each of six different letters. In Experiment 3, subjects learned similar problems. They learned three facts about each of two letters. Thus, when a problem was presented, there were two other problems that shared a letter with it and therefore could interfere with it. In contrast, in Experiments 1 and 2, none of the problems shared letters, so there would be nothing to produce interference.²

Problem frequency was manipulated as in the standard condition of Experiments 1 and 2: Problems with addend 2 were presented 48 times per block, problems with addend 3 were presented 32 times per block, and problems with addend 4 were presented 16 times per block. If differences in frequency produce differences in strength that, in turn, produce differences in the potential to produce in-

terference, then small problems (addend 2) should not suffer much interference from large problems (addend 4) whereas large problems should suffer severe interference from small problems. Whether there are such differences in interference and whether they remain after performance becomes asymptotic was tested in Experiment 3.

Method

Except for the problem sets, the method was the same as in Experiment 2. In Experiment 3, there were nine sets of six problems in which two letters were paired with the digit addends 2, 3, and 4. As in the previous experiments, the different problem sets used 18 letters of the alphabet. The first set used A and B as addends, the second used C and D as addends, and so on up to the last set, which used Q and R as addends. False answers to equations using the first letter addend in a set (e.g., A) were true answers + 1 (e.g., $A + 2 = D$, $A + 3 = E$, $A + 4 = F$), and false answers to equations using the second letter addend in a set (e.g., B) were true answers - 1 (e.g., $B + 2 = C$, $B + 3 = D$, $B + 4 = E$). This was done to ensure that no answer in a subject's set was uniquely true or false. There were 18 subjects selected from the same population sampled in Experiment 2.

Results and Discussion

Mean reaction time and accuracy were computed for each combination of digit addend, truth value, and block for each subject. The means across subjects are presented in Table 3. Accuracy was high, averaging 95.7% and uncorrelated with reaction time [$r = -0.194$, $F(1,88) = 3.43$, $p > .05$]. The average reaction time was 1,287 msec.³

As before, the problem-size effect was estimated as the slope of a straight line fitted to the function relating reaction time to digit addend. The mean slopes in each group are plotted in Figure 2.

The problem-size effect in the first block was not as large as in Experiment 2, but it diminished more slowly and remained above 100 msec until the 10th block. In the last session (Blocks 13-15), it averaged 75 msec, three times the slope in the last session of Experiment 2. These

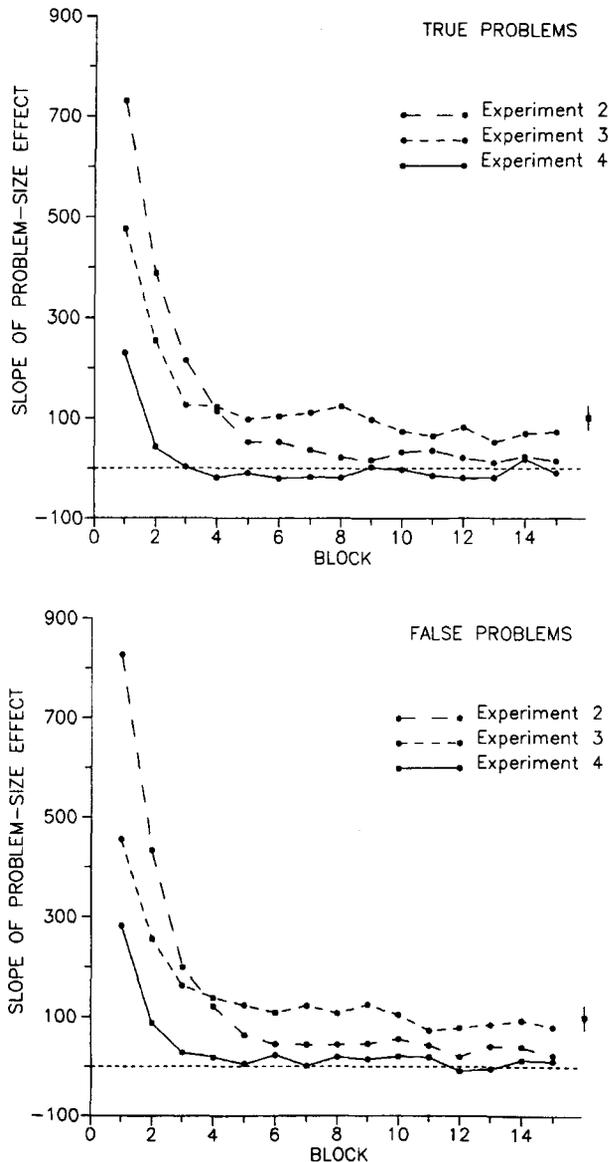


Figure 2. Mean slope of the linear function relating reaction time to digit addend as a function of practice block for true (top panel) and false (bottom panel) problems in Experiments 2 (dashed line), 3 (dotted line), and 4 (solid line). (Error bars for between-experiment comparisons, based on the experiment \times block \times truth error term, are plotted to the right of the data.)

results suggest that there was interference and that differences in frequency resulted in differences in interference, even when performance was at asymptote.

These conclusions were supported in a 15 (block) \times 2 (truth) \times 3 (addend) ANOVA on the mean reaction times and a 15 (block) \times 2 (truth) ANOVA on the slopes. In the reaction time ANOVA, there were significant main effects for block [$F(14,238) = 61.75$, $MS_e = 211,419.76$, $p < .01$], truth [$F(1,17) = 52.99$, $MS_e = 66,380.40$, $p < .01$], and addend [$F(2,34) = 18.72$, $MS_e = 473,233.35$, $p <$

.01] and significant interactions between block and addend [$F(28,476) = 12.12$, $MS_e = 34,341.51$, $p < .01$]. In the slope ANOVA, there was a significant main effect of block [$F(14,238) = 29.04$, $MS_e = 13,012.90$, $p < .01$]. The average slope per block was compared against zero, using the error term from the main effect of block (13,012.90). Every one of the average slopes was significantly greater than zero (LSD = 53, $p < .05$).

Interference was further assessed by comparing the reaction times and slopes with those from Experiment 2. Overall, the average reaction time was 1,287 msec in Experiment 3 and 906 msec in Experiment 2. In Experiment 3, the average reaction time per block was above 1,000 msec in every block and, with the exception of the first block, it was longer than the average reaction time of each corresponding block in Experiment 2 (see Tables 2 and 3). This suggests that, in Experiment 3, there was interference in solving problems that were similar to other problems (i.e., in solving problems that had the same letter addend as two other problems in a set).

This conclusion was confirmed in a 2 (group: Experiment 2 vs. 3) \times 15 (block) \times 2 (truth) \times 3 (addend) ANOVA on the mean reaction times. The main effect of experiments was significant [$F(1,34) = 791.22$, $MS_e = 4,923,839.44$, $p < .01$], and there were significant interactions between experiment and block [$F(14,476) = 6.09$, $MS_e = 330,263.37$, $p < .01$], experiment, block, and truth [$F(14,476) = 4.05$, $MS_e = 15,761.58$, $p < .01$], and experiment, block, and addend [$F(28,952) = 4.28$, $MS_e = 45,957.79$, $p < .01$]. A planned comparison showed that for the last third of the experiment, when performance was close to asymptote (Blocks 11–15), reaction times were significantly longer in Experiment 3 than in Experiment 2 [1,058 vs. 639 msec; $F(1,476) = 143.73$, $MS_e = 330,263.37$, $p < .01$].

Slopes were analyzed in a 2 (group: Experiment 2 vs. Experiment 3) \times 15 (block) \times 2 (truth) ANOVA. The main effect of block was significant [$F(14,476) = 65.64$, $MS_e = 26,219.28$, $p < .01$], and so was the interaction between experiments and blocks [$F(14,476) = 8.06$, $MS_e = 26,219.28$, $p < .01$]. Fisher's LSD test revealed significantly steeper slopes in Experiment 2 than in Experiment 3 for Blocks 1, 2, and 3 and equal slopes for Blocks 4 and 11. However, in the other 10 blocks, slopes were significantly steeper in Experiment 3 (LSD = 36, $p < .05$). A planned comparison showed that the slopes for the last 5 blocks were significantly steeper in Experiment 3 than in Experiment 2 [75 vs. 27 msec; $F(1,476) = 7.73$, $MS_e = 26,219.28$, $p < .01$].

In summary, performance was not as good in Experiment 3 as it was in Experiment 2. Overall, reaction times were slower and the problem-size effect was larger. Performance improved slowly with practice, but at asymptote the problem-size effect was still significantly greater than zero. The results suggest that frequency and interference together may be responsible for the problem-size effect in number arithmetic with adults.

Table 3
Mean Reaction Times (RT, in Milliseconds) and Percentages of Correct Responses (% Corr)
in Experiment 3 as a Function of Block, Addend, and Truth

Block	Addend											
	True						False					
	2		3		4		2		3		4	
RT	% Corr	RT	% Corr	RT	% Corr	RT	% Corr	RT	% Corr	RT	% Corr	
1	1,800	96	2,300	97	2,737	94	1,980	97	2,475	96	2,852	92
2	1,365	97	1,748	96	1,804	98	1,439	98	1,754	96	1,918	93
3	1,277	98	1,518	95	1,467	96	1,336	98	1,673	96	1,567	93
4	1,117	97	1,355	97	1,300	93	1,192	97	1,462	96	1,395	93
5	1,138	96	1,384	96	1,261	94	1,198	98	1,462	96	1,386	93
6	1,113	97	1,375	97	1,238	96	1,193	97	1,461	96	1,326	95
7	984	97	1,214	96	1,132	94	1,049	97	1,263	93	1,226	92
8	1,013	98	1,209	93	1,217	97	1,127	97	1,353	96	1,258	95
9	1,099	97	1,281	97	1,235	94	1,162	97	1,371	97	1,346	95
10	898	97	1,038	95	1,010	94	957	98	1,148	95	1,128	94
11	934	97	1,055	94	1,035	93	999	97	1,157	95	1,104	95
12	966	97	1,114	93	1,094	93	1,069	97	1,204	96	1,202	94
13	894	97	976	95	983	94	956	98	1,106	95	1,093	95
14	917	97	1,052	96	1,025	97	998	98	1,191	95	1,128	97
15	967	97	1,080	95	1,092	95	1,031	98	1,169	95	1,162	96

Experiment 4

Experiment 2 showed that frequency alone cannot account for the problem-size effect in adults, and Experiment 3 suggested that frequency and interference in combination can account for the effect. In Experiment 4, the question was whether interference alone could account for the effect. Subjects were trained with the problems from Experiment 3—three true and three false facts associated with two letters—but the different addends were presented with equal frequency, as in the control condition in Experiment 1. There was as much potential for similarity-based interference in this experiment as there was in Experiment 3, but there was no frequency imbalance to make large problems more prone to interference than small problems. There should be no problem-size effect unless large problems suffered more interference for other reasons (e.g., Campbell & Graham, 1985; Siegler, 1986, 1988).

Method

The method was the same as in Experiment 3. The problem sets were taken from Experiment 3, and each digit addend was presented with equal frequency, as in the control condition of Experiment 1 (i.e., 32 times per block). There were 18 subjects selected from the same population sampled in Experiments 2 and 3.

Results and Discussion

Mean reaction time and accuracy were computed for each combination of digit addend, truth value, and block for each subject. The means across subjects are presented in Table 4. Accuracy was high, averaging 95.7%, and correlated negatively with reaction time [$r = -.028$, $F(1,88) = 7.40$, $p < .01$]. The average reaction time was 1,197 msec.

As before, the problem-size effect was estimated as the slope of a straight line fitted to the function relating reaction time to digit addend. The mean slopes in each group are plotted in Figure 2.

In the first block, the problem-size effect was substantial but not as large as it was in Experiment 3, where frequency was also manipulated. However, in contrast to Experiment 3, where the problem-size remained substantial throughout practice, the problem-size effect in Experiment 4 diminished rapidly to zero by the third block, as in the control condition in Experiment 1 (see Figure 1). From Block 3 on, it remained stable and indistinguishable from zero, averaging 1 msec in the last session (i.e., Blocks 13–15). At asymptote, there was no problem-size effect, and this suggests that interference alone cannot produce the problem-size effect.

The data were analyzed in a 15 (block) \times 2 (truth) \times 3 (addend) ANOVA on the mean reaction times and a 15 (block) \times 2 (truth) ANOVA on the slopes. In the reaction time ANOVA, there were significant main effects of block [$F(14,238) = 48.65$, $MS_e = 195,672.52$, $p < .01$], truth [$F(1,17) = 42.65$, $MS_e = 68,209.94$, $p < .01$], and addend [$F(2,34) = 3.85$, $MS_e = 604,530.20$, $p < .05$] and a significant interaction between block and addend [$F(28,476) = 4.74$, $MS_e = 34,578.66$, $p < .01$]. In the slope ANOVA, there was a significant main effect of block [$F(14,238) = 9.28$, $MS_e = 17,415.06$, $p < .01$]. Slopes were compared against zero, using the error term from the main effect of block (LSD = 61; $MS_e = 17,415.06$, $p < .05$). Except for Blocks 1 and 2, average slopes were not significantly different from zero. This supports the conclusion that at asymptote, the amount of interference in solving large and small problems was equivalent when their frequencies were equivalent.

Experiment 4 was compared with Experiment 3 in a 2 (group: Experiment 4 vs. Experiment 3) \times 2 (truth) \times 3 (addend) ANOVA on the mean reaction times and in a 2 (group: Experiment 4 vs. Experiment 3) \times 15 (block) ANOVA on the slopes. In the reaction time ANOVA, there were significant main effects of block [$F(14,476) = 109.61$, $MS_e = 203,546.14$, $p < .01$], truth [$F(1,34) = 95.28$,

Table 4
Mean Reaction Times (RT, in Milliseconds) and Percentages of Correct Responses (% Corr)
in Experiment 4 as a Function of Block, Addend, and Truth

Block	Addend											
	True						False					
	2		3		4		2		3		4	
RT	% Corr	RT	% Corr	RT	% Corr	RT	% Corr	RT	% Corr	RT	% Corr	
1	1,870	95	2,107	94	2,327	93	1,901	95	2,352	96	2,468	93
2	1,342	97	1,493	96	1,426	96	1,387	96	1,579	96	1,556	95
3	1,209	98	1,295	97	1,213	95	1,283	97	1,454	95	1,340	95
4	1,136	97	1,254	98	1,099	96	1,200	97	1,347	96	1,238	96
5	1,079	97	1,188	97	1,058	98	1,138	96	1,234	95	1,147	96
6	1,089	97	1,157	95	1,051	95	1,102	97	1,251	96	1,147	96
7	1,055	96	1,135	94	1,019	96	1,119	97	1,241	95	1,118	95
8	1,096	95	1,160	97	1,060	97	1,115	98	1,266	96	1,147	96
9	1,087	98	1,176	96	1,089	96	1,107	96	1,219	94	1,139	95
10	912	95	1,033	97	909	96	975	96	1,115	94	1,018	97
11	949	96	1,027	94	919	94	1,006	97	1,140	95	1,041	96
12	1,004	95	1,066	96	966	97	1,051	94	1,180	94	1,035	94
13	967	97	1,015	96	931	96	1,010	96	1,128	95	1,000	94
14	952	95	1,070	96	990	96	1,016	97	1,137	94	1,040	94
15	962	95	1,070	95	939	95	1,051	97	1,155	94	1,072	96

$MS_e = 67,295.17, p < .01$], and addend [$F(2,68) = 16.95, MS_e = 538,881.77, p < .01$]. The main effect of experiment was not significant [$F(1,34) < 1, MS_e = 7,802,850.97$]. Overall, mean reaction times were similar in Experiment 3 (i.e., 1,287 msec) and Experiment 4 (i.e., 1,197 msec) and using the error term from the block \times experiment interaction (203,546.14), Fisher's LSD test revealed that average reaction times per block did not differ significantly between experiments in Blocks 7, 8, 10, 11, 12, 13, 14, and 15 (LSD = 86, $p < .05$). The similar and longer reaction times in Experiments 3 and 4, relative to those in Experiment 2 (where problems were never similar to other problems), suggests that similarity-based interference was operating in both Experiments 3 and 4. However, the effect of addend was larger in Experiment 3 than in Experiment 4, producing a significant interaction between addend and experiment [$F(2,68) = 3.80, MS_e = 538,881.77, p < .05$], which supports the conclusion that the amount of similarity-based interference was equivalent in large and small problems in Experiment 4 but not in Experiment 3.

In the slope ANOVA, there were significant main effects of experiment [$F(1,34) = 16.39, MS_e = 206,255.65, p < .01$], block [$F(14,476) = 33.38, MS_e = 15,213.98, p < .01$], and truth [$F(1,34) = 4.47, MS_e = 22,846.72, p < .05$] and a significant interaction between experiment and block [$F(14,476) = 2.08, MS_e = 15,213.98, p < .05$]. A planned comparison showed that for the last third of the experiments when performance was at asymptote (Blocks 11–15) slopes were significantly larger in Experiment 3 than in Experiment 4 [$F(1,476) = 34.22, MS_e = 15,213.98, p < .01$]. These comparisons, together with those between Experiments 2 and 3, suggest that neither interference alone nor frequency alone can account for the problem-size effect at asymptote but that a combination of interference and frequency can.

GENERAL DISCUSSION

Why is there a problem-size effect in adults? The present experiments compared the strength hypothesis with the hypothesis that a combination of strength and interference is responsible for the effect. Experiment 1 showed that reaction time depended more on the frequency with which problems were presented than on the size of the problems. This suggests that strength may be sufficient to explain the problem-size effect at low levels of practice. When small problems were more frequent than large problems (as is the case in children's arithmetic texts), the problem-size effect was exaggerated and remained well above zero even at the end of practice.

Experiment 2 asked whether the problem-size effect would remain even at asymptotic levels of practice. Subjects were given five times as much practice as in Experiment 1, and the problem-size effect approached zero well before the end of practice. Differences in strength (as a function of differences in frequency) no longer had an effect. This suggests that strength may not be sufficient to explain the problem-size effect at asymptotic levels of performance. It may be sufficient to explain it at low levels of practice but not at high levels of practice.

Experiment 3 repeated the procedure of Experiment 2 but increased the similarity of the problems subjects learned. Presumably, the greater the similarity, the greater the potential for interference. Although the problem-size effect diminished with practice, it remained significantly greater than zero even at asymptote. By contrast, in Experiment 2, there was less potential for interference (problems were dissimilar) and there was no problem-size effect with extended practice. This suggests that a combination of strength and interference may explain the problem-size effect at high levels of practice.

Experiment 4 used the problem sets from Experiment 3, in which the problems were similar to each other, but presented different problems equally often, as in the control condition of Experiment 1. With this design, interference should be equal for large and small problems, and there should be no problem-size effect. The finding that there was no problem-size effect suggests that similarity-based interference alone may not explain adult arithmetic performance.

In the present experiments, interference was conceptualized as competition between problems that were similar on the basis of a shared addend. However, there are other types of interference to which large problems may be more prone than small problems. For example, Siegler (1986, 1988) argued that more errors are made in solving large problems than in solving small problems, so large problems have more error responses associated with them. These error responses are activated and interfere with correct associations each time problems are presented. Campbell and Graham (1985) argued that large problems might be more prone to interference than small problems because they are often learned later than small problems (Campbell & Graham, 1985). Small problems are learned in a context in which there are few competing associations (i.e., they are learned first), whereas large problems are learned in a context in which there are strong competing associations (from small problems). Large problems suffer from more proactive interference than do small problems, and are therefore not learned as well as small problems. Campbell and Oliphant (1992) argued that large problems might be prone to more interference than would small problems because answers to large problems are less discriminable from answers to their large-magnitude neighbors than answers to small problems are from answers to their small-magnitude neighbors (e.g., the difference between the answers 63 and 64 is psychologically smaller than the difference between the answers 8 and 9, in solving the problem 9×7 vs. 2×4). The lower the discriminability, the more interference there is from neighboring competing answers. Thus, large problems suffer more interference than small problems.

It is not clear whether any type of interference alone can account for the problem-size effect at asymptote. The present results suggest that at least similarity-based interference alone cannot. However, they also suggest that a combination of strength and similarity-based interference can. Given differences in strength (via frequency) and potential for interference (via similarity), stronger small problems interfered more with weaker large problems than vice versa, even after extended practice. It may be that as long as there are differences in strength, stronger problems will interfere more with weaker problems than vice versa, regardless of the type of interference that is operating.

A few researchers have explored interference in arithmetic in simulation models of arithmetic (e.g., Campbell & Oliphant, 1992; Clark, 1992; McCloskey & Cohen, 1989). However, the precise delineation of the mechanisms of interference, how they are related to strength, and how they interact with skill and practice in human arithmetic performance awaits further research.

Alphabet-arithmetic tasks may be useful for investigating and discriminating accounts of interference in human subjects. For example, the idea that some problems suffer from more proactive interference than others could be tested by introducing subjects with some alphabet-arithmetic problems before others in a training sequence and seeing if training order made a difference. Moreover, by manipulating both training order and frequency, one could assess the effects of training order relative to the effects of frequency. The idea that large problems are less discriminable than small problems could be tested by assessing interference within a group of small problems (e.g., $A + 2$, $A + 3$, $A + 4$) relative to the interference within a group of large problems (e.g., $A + 7$, $A + 8$, $A + 9$). Such ideas have been implemented in computer models of numerical arithmetic (e.g., Campbell & Oliphant, 1992; McCloskey & Lindemann, 1992), but it would be useful to know the extent to which the determinants of simulated performance reflect the determinants of subjects' performance under the same manipulations. With numerical arithmetic tasks, such manipulations may not be feasible because they would be confounded by subjects' previous experience. Except for children who are about to learn arithmetic, everyone has had some experience with problems of certain sizes, frequencies, and training orders which they would bring into the laboratory. Moreover, manipulating such variables may not be feasible in numerical arithmetic because of ethical considerations. One manipulation may produce better learning than another, and it may not be ethical to place subjects in conditions that may result in less than optimal learning. With alphabet arithmetic, subjects do not have any previous experience to contaminate results and there are no ethical implications of being in one condition or another.

What we can learn from alphabet arithmetic is useful inasmuch as it can inform us about what happens in numerical arithmetic. This depends on the generalizability of alphabet arithmetic to numerical arithmetic. Unfortunately, alphabet arithmetic may not mimic the "ecology" of numerical arithmetic very well. Unlike the laboratory situations in which adults solve alphabet-arithmetic problems for a short period (from 15 min to 16 h; see Logan & Klapp, 1991), numerical arithmetic learning begins in childhood and continues for many years (for at least 12 school years). However, alphabet arithmetic does mimic the theoretical processes that underlie numerical arithmetic quite well. As with numerical arithmetic, there is evidence that initially problems are solved by counting and that with practice there is a transition to remembering which then dominates asymptotic performance. Thus, alphabet arithmetic may be a good task to use in assessing the validity of explanations based on counting and remembering processes.

REFERENCES

- ASHCRAFT, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, *44*, 75-106.
- ASHCRAFT, M. H., & BATTAGLIA, J. (1978). Cognitive arithmetic: Evidence for retrieval and decision processes in mental addition. *Journal of Experimental Psychology: Human Learning & Memory*, *4*, 527-538.
- ASHCRAFT, M. H., & CHRISTY, K. S. (in press). The frequency of arith-

- metic facts in elementary texts: Addition and multiplication in grades 1-6. *Journal for Research in Mathematics Education*.
- BENFORD, F. (1938). The law of anomalous numbers. *Proceedings of the American Philosophical Society*, **78**, 551-572.
- CAMPBELL, J. I. D. (1987). Network interference and mental multiplication. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 109-123.
- CAMPBELL, J. I. D. (1990). Retrieval inhibition and interference in cognitive arithmetic. *Canadian Journal of Psychology*, **44**, 445-464.
- CAMPBELL, J. I. D. (1991). Conditions of error priming in number-fact retrieval. *Memory & Cognition*, **19**, 197-209.
- CAMPBELL, J. I. D., & CLARK, J. M. (1989). Time course of error-priming in number fact retrieval: Evidence for excitatory and inhibitory mechanisms. *Journal of Experimental Psychology: General*, **117**, 204-214.
- CAMPBELL, J. I. D., & GRAHAM, D. J. (1985). Mental multiplication skill: structure, process, and acquisition. *Canadian Journal of Psychology*, **39**, 338-366.
- CAMPBELL, J. I. D., & OLIPHANT, M. (1992). Representation and retrieval of arithmetic facts: A network-interference model and simulation. In J. I. D. Campbell (Ed.), *The nature and origins of mathematical skills* (pp. 331-364). Amsterdam: Elsevier, North-Holland.
- CLAPP, F. L. (1924). *The number combinations: Their relative difficulty and the frequency of their appearance in textbooks* (Bulletin No. 2). University of Wisconsin Bureau of Educational Research.
- CLARK, J. M. (1992). Inhibitory mechanisms in normal and dysfunctional number processing. In J. I. D. Campbell (Ed.), *The nature and origins of mathematical skills* (pp. 411-456). Amsterdam: Elsevier, North-Holland.
- COMPTON, B. J., & LOGAN, G. D. (1991). The transition from algorithm to retrieval in memory-based theories of automaticity. *Memory & Cognition*, **19**, 151-158.
- GEARY, D. C., WIDAMAN, K. F., & LITTLE, T. D. (1986). Cognitive addition and multiplication: Evidence for a single memory network. *Memory & Cognition*, **14**, 478-487.
- GROEN, G. J., & PARKMAN, J. M. (1972). A chronometric analysis of simple addition. *Psychological Review*, **79**, 329-343.
- GROEN, G. J., & RESNICK, L. B. (1977). Can preschool children invent addition algorithms? *Journal of Educational Psychology*, **69**, 645-652.
- HAMANN, M. S., & ASHCRAFT, M. H. (1986). Textbook presentations of the basic addition facts. *Cognition & Instruction*, **3**, 173-192.
- LOGAN, G. D., & KLAPP, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 179-195.
- MCCLOSKEY, M., & COHEN, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 24, pp. 109-165). San Diego: Academic Press.
- MCCLOSKEY, M., & LINDEMANN, A. M. (1992). Mathnet: Preliminary results from a distributed model of arithmetic fact retrieval. In J. I. D. Campbell (Ed.), *The nature and origins of mathematical skills* (pp. 365-409). Amsterdam: Elsevier, North-Holland.
- SIEGLER, R. S. (1986). Unities across domains in children's strategy choices. In M. Perlmutter (Ed.), *Minnesota Symposium on Child Psychology* (Vol. 19, pp. 1-48). Hillsdale, NJ: Erlbaum.
- SIEGLER, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, **116**, 250-264.
- SIEGLER, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General*, **117**, 258-275.
- SIEGLER, R. S., & SHRAGER, J. (1984). A model of strategy choice. In C. Sophian (Ed.), *Origins of cognitive skills*. Hillsdale, NJ: Erlbaum.
- SVENSON, O. (1985). Memory retrieval of answers of simple additions as reflected in response latencies. *Acta Psychologica*, **59**, 285-304.
- THOMAS, H. B. G. (1963). Communication theory and the constellation hypothesis of calculation. *Quarterly Journal of Experimental Psychology*, **15**, 173-191.
- THORNDIKE, E. L. (1922). *The psychology of arithmetic*. New York: Macmillan.
- WINKELMAN, J. H., & SCHMIDT, J. (1974). Associative confusions in mental arithmetic. *Journal of Experimental Psychology*, **102**, 734-736.
- ZBRODOFF, N. J. (1979). *Development of counting and remembering as strategies for performing simple arithmetic in elementary school children*. Unpublished master's thesis, University of Toronto.
- ZBRODOFF, N. J., & LOGAN, G. D. (1986). On the autonomy of mental processes: A case study of arithmetic. *Journal of Experimental Psychology: General*, **115**, 118-130.
- ZBRODOFF, N. J., & LOGAN, G. D. (1990). On the relation between production and verification tasks in the psychology of simple arithmetic. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 83-97.

NOTES

1. Note that in Siegler and Shrager's (1984) model, arithmetic performance depends on a combination of interference and strength factors. In particular, the probability of retrieving a correct answer is a function of the strength of correct associations relative to the strength of incorrect associations.

2. In all experiments, interference could be produced via associations with the digit addend. Each digit occurred with at least two problems, and the different answers associated with the digit could be retrieved and produce interference. However, the frequency with which the digits are presented was the same in Experiments 2 and 3, so differences in performance should be attributed to interference produced by the letters rather than to interference produced by the digits.

3. As of the third block in Experiment 3 (and the second block in Experiment 4), there was a quadratic trend in the mean reaction times (also see Compton & Logan, 1991). The linear increase in reaction time flattened out for the largest addend (i.e., 5). This reflects a transition from counting to retrieval. The transition occurs sooner for larger problems because the counting process is harder and takes longer. Some subjects reported that they deliberately tried to memorize answers to larger problems in order to avoid counting. Also, in a race model like Compton and Logan's (1991), in which counting processes race against retrieval processes and counting processes are stable over practice whereas retrieval processes speed up with practice, the larger problems, which take longer to count, will be the first to lose the race. Thus, the larger problems will flatten out sooner.

(Manuscript received July 19, 1994;
revision accepted for publication September 13, 1994.)