Mental addition: A test of three verification models

MARK H. ASHCRAFT and EDMUND H. STAZYK
Cleveland State University, Cleveland, Ohio 44115

Three explanations of adults' mental addition performance, a counting-based model, a direct-access model with a backup counting procedure, and a network retrieval model, were tested. Whereas important predictions of the two counting models were not upheld, reaction times (RTs) to simple addition problems were consistent with the network retrieval model. RT both increased with problem size and was progressively attenuated to false stimuli as the split (numerical difference between the false and correct sums) increased. For large problems, the extreme level of split (13) yielded an RT advantage for false over true problems, suggestive of a global evaluation process operating in parallel with retrieval. RTs to the more complex addition problems in Experiment 2 exhibited a similar pattern of significance and, in regression analyses, demonstrated that complex addition (e.g., $14 + 12 = 26$) involves retrieval of the simple addition components ($4 + 2 = 6$). The network retrieval/decision model is discussed in terms of its fit to the present data, and predictions concerning priming facilitation and inhibition are specified. The similarities between mental arithmetic results and the areas of semantic memory and mental comparisons indicate both the usefulness of the network approach to mental arithmetic and the usefulness of mental arithmetic to cognitive psychology.

Despite the pervasiveness of mental arithmetic in adult cognition, little agreement can be found within cognitive psychology as to the memory structures and processes that underlie this performance. The models that have been proposed for mental arithmetic merely emphasize this lack of agreement; they range from purely declarative or prestorage to purely computational or counting explanations. The purpose of this research was first to test these models in the domain of simple mental addition and then to assess their generality when slightly more complex mental addition is required. A basic criterion in this evaluation is that any serious model should not only account for results in mental addition but also be compatible in form with other arithmetic/number operations that adults perform mentally.

The result of primary interest in mental addition research is termed the "problem-size effect": Reaction time (RT) to positives increases as the size of the problem increases. Possibly the simplest explanation of this effect, and the simplest definition of problem size, is embodied in Groen and Parkman's (1972) "min" (for minimum addend) model. According to this model, mental addition is performed by setting an internal counter to the larger (maximum) of the two addends in a simple addition problem and then incrementing this value, by ones, the number of times specified by the smaller (minimum) addend. Since counter-setting time

should be constant (but cf. Banks, Fujii, & Kayra-Stuart, 1976), increases in RT should be proportional to the number of increments added, that is, to min. More formally, this model predicts that $RT = c + \text{positive } f(\text{min})$, where c is a combination of encoding, counter-setting, decision, and response execution times, and the positive function of min is linear.

Groen and Parkman's (1972) data clearly supported this model for first-graders tested in a production task. In Groen and Parkman's regression analysis, the min model accounted for 80% of the variance in RT, whereas no other model accounted for more than 32%. In a true-false verification study with adults, however, Parkman and Groen (1971) found that the sum of the problem was almost as good a predictor as min (71% vs. 73%, respectively).

While the Groen and Parkman (1972) counting model could easily be revised to use sum rather than min in the incrementing process (set counter to zero, increment by one addend and then the other), Groen and Parkman nonetheless doubted the generality of either a min- or a sum-based counting model for adults. Primary among their reasons was that incrementing time in the first-graders was estimated to be 400 msec/increment, but only 20 msec/increment for adults. Since incrementing in this model was viewed as quite similar to "silent counting," an adult rate of 20 msec seemed very unlikely (although the possibility of a fast incrementing process of unknown nature was left open). A second reason for Groen and Parkman's doubts involved RT to ties, that is, to problems in which a number is added to itself. In both their child and adult samples, RT to ties appeared to be a constant value, unaffected by the problem-

size variables that seemed important for nontie problems. As an explanation for this effect, Groen and Parkman proposed that ties were exceptions to the rule, that is, that they were retrieved from some "fast-access" memory store.

Three further objections to the min model should be noted here. First, evidence from a study by Winkelman and Schmidt (1974) suggests strongly that addition and multiplication are stored in interrelated fashion in memory (Ashcraft & Battaglia, 1978, found suggestive evidence of a similar interrelationship between addition and subtraction). Such interrelationships are virtually precluded in a computation/counting-based model, however. Second, and regardless of the developmental results, counting models as a class provide exceptionally awkward explanations of other simple arithmetic operations, such as multiplication (see also Parkman, 1972). Finally, Ashcraft and Battaglia (1978) found in two experiments that minimum addend was a less successful predictor of adults' RTs than the square of the problem's sum. While no special significance attaches to the squared power per se, the importance of this effect was that an exponentially increasing RT function is quite difficult to reconcile with an incrementing-based process.

A more complex model, also suggested by Groen and Parkman (1972), meets some of these criticisms. The model was derived from the fast-access notion developed for tie problems. According to this revised model, most of adult performance should be viewed as the outcome of fast- or direct-access retrieval from memory, such retrieval requiring an essentially constant amount of time (i.e., a regression slope of zero). As such, the 20-msec slope across min was viewed as merely an average of the many fast-access trials with those few very slow counting trials on which direct access had failed. In their data, direct access was estimated to fail on about 5% of the trials, necessitating the backup counting process, which proceeds at the child-like rate of 400 msec (probability of retrieval failure = 20 msec/ 400 msec = .05). Formally, then, the direct-access model predicts that overall RT will be the average of two equations, $RT_1 = c$, with probability of .95, and $RT_2 = c +$ positive f(min), with probability of .05. In these equations, c is the combination of encoding, direct-access, decision, and response execution times; the positive function of min is still a 400-msec linear slope.

Given a few generous assumptions, this direct-access/ counting model can handle two of the three objections noted above with fair success. First, the direct-access notion requires that some memory structure be available for retrieval of addition answers, although Groen and Parkman (1972) did not speculate about this structure (but see Parkman, 1972). Given that a structure must be present, it is probably not unreasonable to assume further that other arithmetic operations would be similarly represented in memory and that these representations would be interconnected in some relevant fashion. This scheme could provide a framework within

which the confusion/interference effects found by Winkelman and Schmidt (1974) could be explained (although it is not entirely clear how direct access to one portion of the representation could be influenced by the interconnections in the representation itself). Second, the direct-access model relies very little on a counting-based operation for adult performance, possibly as little as 5%. As such, the general difficulties with counting models are at least minimized. Substitution of a different backup process, say reattempted retrieval, could preserve the direct-access component and achieve generality to other operations simultaneously.

The final objection remains somewhat of a problem for the direct-access model, however. The direct-access model in fact makes the same RT predictions as the simpler min model, a linear increase in RT as a function of minimum addend. Thus Ashcraft and Battaglia's (1978) finding of an exponential increase seems to disconfirm a major prediction from the Groen and Parkman (1972) direct-access model. (Unfortunately, it is not possible to determine if there was an exponential trend in Groen and Parkman's data, since they did not include such a predictor variable in their analyses and since their RTs were averaged across 4 days of practice.)

As a consequence of this disconfirmation, Ashcraft and Battaglia (1978) proposed a four-stage retrieval/ decision model, in which time-consuming search through a network representation accounts for the observed RT increase to positives. This proposal was in fact quite similar to one suggested by Parkman (1972) in a discussion of multiplication. In both schemes, the mental representation was hypothesized to be functionally organized as a printed addition table, in which the sum of two numbers is located at the intersection of the appropriate column and row. After encoding of the problem, retrieval of the sum from the network occurs, with elapsed retrieval time proportional to the distance traversed during the search. Ashcraft and Battaglia further assumed that the mental table was "stretched" in the region of larger sums, thus attributing (in post hoc fashion) the exponentiality in RT to greater subjective distance in the network structure. A decision stage follows retrieval, during which the result of the memory search, the sum of the problem, is compared with the answer presented in the stimulus, thus enabling the yes-no reponse. Decision time for true problems was assumed to be essentially constant and quite rapid. Thus for positives, RT = c + positive f(problem size), where c is the combination of encoding, decision, and response execution times, and the positive function of problem size appears to be the exponentially increasing function of sum squared. For problems presented with an incorrect answer, however, the comparison of the retrieved sum with the stated incorrect sum results in an inequality judgment. Decision time in this situation is therefore inversely proportional to the degree of mismatch or "split" (the difference between correct and incorrect), an RT effect found in addition by Ashcraft and Battaglia

and in the simpler inequality judgment paradigm used by Banks and others (e.g., Banks et al., 1976). Thus, for negatives, RT is both a positive function of problem size and a negative function of split, with these functions related to retrieval and decision, respectively. Note that such a decision stage is appropriate only to the verification performance under consideration here; it would not play a role in production tasks.

Regardless of the advantages of this retrieval/decision model, a direct-access explanation (based on Groen & Parkman, 1972) of Ashcraft and Battaglia's (1978) results is in fact available, although it requires a modification of Groen and Parkman's model. Rather than assume that the probability of retrieval failure varies as "a function of some structural variable other than the min," assume instead that retrieval failure increases proportionally with increases in min. Since the counting backup process will be of longer duration as problem size (min) increases, the additional effect of more frequent retrieval failure at large values of min could easily yield RTs that are best fit by an exponential curve. Even the substitute process of reattempted direct access might produce the same exponentiality, given a positive relationship between problem size and retrieval failure.

Therefore, the major purpose of Experiment 1 was a test of these rival explanations of the problem-size effect, direct-access retrieval vs. time-consuming search through a network representation. If the data support direct access, then attention must be devoted both to the mechanisms that operate after retrieval failure as well as to decision stage processes implied by the split effect. Alternatively, if evidence against direct-access processing is obtained, then attention can be directed toward further specification of the Ashcraft and Battaglia (1978) network model. We include both problem size and split here to elaborate a process model of verification performance in either situation.

EXPERIMENT 1

Method

Twenty undergraduates enrolled at the Cleveland State University served as subjects in the 45-min sessions. All subjects were given credit in introductory psychology courses for their participation.

A total of 200 addition problems, presented in column form, served as stimuli. Of these, 100 were the true "basic 100 additions facts," that is, the pairs 0 + 0 up to 9 + 9, along with the correct sum. For the 100 false stimuli, the same addition problems were assigned an incorrect answer, in which the difference between correct and incorrect (the split) was ±1, 5, 9, or 13. Incorrect answers came from the same range as the correct sums, namely, 0 through 18, and to the extent possible mirrored the frequency of correct sums. Given the range restriction, notice that it is not possible to assign problems to the Split 13 condition in an entirely random fashion. That is, a problem such as 5 + 4 cannot appear in the Split 13 condition, since the possible incorrect answers (−4 or 22) are out of the admissible range. As a result, the split 13 condition contained either small or large problems, but none in the middle range of size. Despite this, the Split 13 condition was deemed important enough to be included in the study, since Ashcraft and Battaglia (1978) proposed that extreme splits might result in processing that is very

different from, possibly parallel to, the processing that is normally obtained. These Split 13 data, however, were not included in the major analysis of variance due to the nonrandom assignment of stimuli. Adjunct analyses, therefore, will be reported for these data.

The 200 stimuli, preceded by 20 practice trials, were sequenced randomly, with the constraints that no repetition of numbers across consecutive problems be allowed and no more than four consecutive trials require the same response. Stimuli were rear-projected onto a screen approximately 75 cm from the subject. An electronic clock timer was activated at stimulus presentation and was stopped by the subjects' responses to either the "true" or the "false" button. Half of the subjects responded "true" to the right button, and half to the left. A 4-sec blank interval followed each response.

Results and Discussion

Overall error rate for the 20 subjects was 4.1%, and extreme scores (for Dixon's test; see Wike, 1971) accounted for only 2.7% of the data points. Errors and extreme scores were positively related to RT, suggesting no limitations on interpretation due to a speed-accuracy tradeoff.

The design of the analysis of variance included three within-subjects factors, true-false, split (±1, 5, 9), and problem size (small vs. large; i.e., sums 0-9 vs. 10-18). Each subject's mean RT to the appropriate set of problems served as the dependent variable. Two points should be made about this design (and that of Experiment 2). First, split is by definition manipulated only in false stimuli, since only in false stimuli is there a difference between the correct and stated sums. While the present design yokes each false problem (e.g., 4 + 5 = 7) to its corresponding true version (4 + 5 = 9), no split effect is possible, of course, on true problems. Thus the paradoxical "true/small-split" condition in the figures, for example, merely refers to those true problems (4 + 5 = 9) that appeared with a small split in the false stimulus set. Second, the separation of problems into small vs. large is admittedly somewhat crude (although probably no more so than similar dichotomous manipulations in semantic memory research, for example). On the other hand, information concerning interactions between problem size and other factors is more conveniently represented and understood in the analysis of variance framework. Further, we rely on the results of regression analyses for detailed information about the problem-size effect per se.

It seems appropriate at this point to mention the procedures and advice that guided the regression analyses reported here, as well as our definitions and criteria of significance. We follow the terminology and suggestions of Draper and Smith (1966) and Myers (1979) in selecting "the best regression equation" for the present data, that is, data involving a set of highly intercorrelated predictor variables, only two of which have been advanced as central to theory. In the absense of well-developed theory, Draper and Smith recommend the forward stepwise regression procedure, in which variables are (1) selected one by one on the basis of the variables' predictive relationship to unaccounted for

variance and (2) reexamined after each additional variable is selected, to insure the continued significance of each individual predictor. Any variable that is rendered nonsignificant at a later point in the regression is removed from the model (in fact, this situation was never encountered in the present results). Such a procedure necessitates three classes of F tests for each regression analysis. First, the normal test of the equation's significance is required, F_{reg}. Second, a variable's contribution to the prediction equation must be tested before it is entered into the model. This test is referred to as "F-to-enter" or "sequential F," and it tests the null hypothesis that the increment in variance accounted for by the added variable is zero.[1] Finally, each variable currently in a regression equation must be tested as to its independent significance; this partial F test is the basis for either keeping or removing a variable from an equation. All three classes of F tests are reported here, with the understanding that the partial F value for a variable that entered the equation at that step is identical to that variable's F-to-enter (see Draper & Smith, 1966).

Having specified our methods and criteria, it must be admitted that this does not truly ameliorate the problem of intercorrelations among the predictor variables. Such multicollinearity, particularly with a large number of predictors, may lead to several difficulties, including spurious significance, unreliable estimates of coefficients, and different orders of inclusion when new samples are tested. We would argue that these potential problems are in fact probably not severe in these data, for the following reasons. First, three independent samples of adult subjects have been tested (two in Ashcraft & Battaglia, 1978; one here in Experiment 1), with remarkably similar regression results, in terms of both inclusion orders and the actual values of coefficients (the inclusion orders have been replicated in fourth- and sixth-grade samples as well; see Ashcraft & Fierman, Note 1). Second, the subsets of predictor variables behave in a very consistent fashion; for example, one of the problem-size variables has always entered the equation first, and apparently trivial variables (odd-even, first and second addends) have almost never played a significant role in any equation. Third, Experiment 2 here demonstrates consistency and generality of results when new samples of both subjects and stimuli are tested. These reasons suggest that interpretations of the present results may not be as tenuous as would otherwise be expected. Finally, a major goal of the present research is to specify a theoretical approach to mental arithmetic, specifically a model that will dictate the a priori basis for selecting and evaluating predictor variables in future research.

In the analysis of variance, all main effects and two-way interactions were significant beyond the .01 level. Briefly, true problems were 100 msec faster than false problems [$F(1,19) = 35.03$, MSe = 17,486.1], and small problems were 175 msec faster than large problems [$F(1,19) = 49.53$, MSe = 36,628.5]. This problem-size

effect was magnified to 210 msec for true problems, but it declined to 130 msec for false problems [for the interaction, $F(1,19) = 10.59$, MSe = 8,420.4].

The main effect of split [$F(2,38) = 39.99$, MSe = 10,303.7] was qualified by its interaction with true-false [$F(2,38) = 17.01$, MSe = 7,821.7]. The interaction revealed a significant RT decline for false problems as split increased (from 1,247 msec to 1,040 msec), but, of course, there was no change in true RT as a function of split. Further, the Split by Problem-Size interaction was significant [$F(2,38) = 9.20$, MSe = 7,908.3]. At Split 1, RT increased from 1,030 msec to 1,275 msec for small to large problems. At Splits 5 and 9, this pattern was attenuated, from 960 msec to approximately 1,100 msec for small to large problems. Splits 5 and 9 did not differ from each other, but both differed from Split 1.

Matched t tests (all conducted at the .01 level to adjust the experimentwise error rate) were performed to pursue the suggestion by Ashcraft and Battaglia (1978) that extreme levels of split might lead to parallel operation of retrieval and decision stages. Figure 1 presents the following RT data for the different conditions. For small problems, mean RTs were significantly slower to false problems at the lower levels of split (all ps < .001); at Split 13, true and false RTs were essentially equivalent (921 vs. 918 msec). For large problems, however, false RTs at Split 13 were over 200 msec faster than true RTs (t = 3.85, p < .01). In contrast, no RT differences were obtained at the Split 5 or 9 levels here, although at Split 1 the false problems were again significantly slower than trues.

In order to examine more detailed aspects of performance, in particular the problem-size effect, mean RT was calculated for each of the 200 stimulus problems and was used as the dependent variable in several forward stepwise multiple regressions. The same set of predictor variable used by Ashcraft and Battaglia (1978) was employed here as well. Briefly, these variables were as follows (starred variables apply to false stimuli only): numerical values of the first and second addends, of the
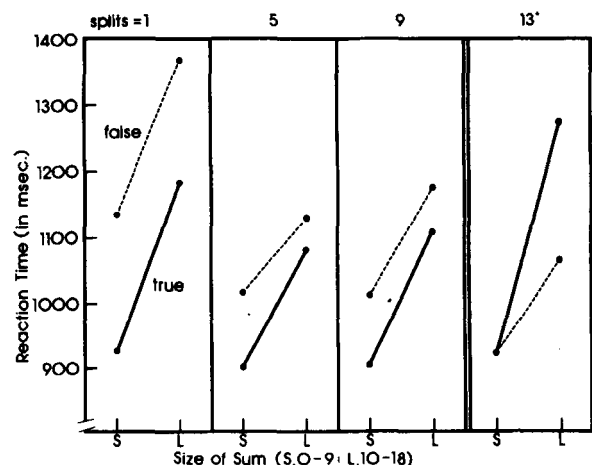

Figure 1. Mean reaction time (RT) to true and false problems as a function of size of sum (S = small, 0-9, L = large, 10-18) and split. *Analyzed separately.

minimum and maximum addends, of the correct sum and *stated sum, of the *difference between correct and stated sums as well as the *absolute value of this difference (the split), and of the square of the problem's sum. Dummy variables were used to represent true vs. false, one- vs. two-digit sum, and the odd/evenness of the two addends, correct sum, and *stated sum.

Mean RT to ties, that is, to problems in which a number is added to itself, was considered first. In this analysis, no significant linear relationship between RT and any considered variable was found. This is a consistent finding in this research, that RT to ties remains constant regardless of problem size. The result suggests that ties enjoy a high degree of accessibility in memory, much more so than nontie problems, or, under a direct-access model, that ties are uniformly processed via direct access, whereas other problems occasionally experience retrieval failure. Interestingly, RT to false ties was not invariant but varied inversely with split. While this is a potentially important result, it is based on too few cases here to support any major interpretation.

The first major regression analysis involved RTs to true, nontie problems (mean = 1,039 msec). Table 1 presents the correlation matrix from this analysis. As found in previous research, the problem-size variables of correct sum, correct sum squared, and minimum addend were the variables most strongly related to RT. In the regression, the slight difference in correlation between RT and min/correct sum squared (.777 vs. .775) led to the inclusion of min on the first step (see Table 2 for statistical summaries of the regression analyses). Despite their high degree of overlap (intercorrelation = .889), correct sum squared was the next significant variable to enter the equation. In this two-variable prediction equation, each variable independently contributes a significant amount of predictive accuracy, and no other variable in the considered set improved the equation significantly. Notice that this pattern was almost exactly repeated when correct sum squared was forced as the first variable: on the first step, correct sum squared accounted for 60% of the variance; then minimum addend was selected on Step 2, increasing the variance

accounted for to 63%. In either case, the raw-score prediction formula in milliseconds was RT = 823 + 39.2 (min) + 1.17 (correct sum squared). Multiple R was .796, and the standard error of the estimate was 127.8.

For the false, nontie problems (mean = 1,094 msec), the history of the best regression solution was as follows: Minimum addend, with a slightly higher correlation (.446) than correct sum (.439) or split (−.445), was selected on the first step. On Step 2, split was selected, joined by the stated but incorrect sum on the third step. The three-term equation was significant and accounted for 55% of the variance in RT (see Table 2). These results replicate the Ashcraft and Battaglia (1978) results on false problems in the general sense that a problem-size variable, correct sum in Ashcraft and Battaglia and min here, then the manipulated split factor, and then the stated sum were selected.

In general, then, RT to both true and false problems is quite sensitive to problem size, regardless of the precise nature of the size variable (min, sum, etc.). This suggests, of course, that at least some of the same processes occur for both types of stimuli, although the nature of those processes is disputed. The following section discusses aspects of the data that are especially relevant as tests of the direct access vs. network retrieval models. Equally important as the problem-size effect for this discussion is the significant influence of split and stated sum on RT to false problems. Such a result strongly indicates the operation of a decision-like stage in verification performance, and this possibility is also explored below.

Specific Tests of the Models

Since Groen and Parkman (1972) themselves suggested that the min counting model was probably inappropriate for adults, only brief mention of the present disconfirming results will be made. In the regression of true, nontie problems, both minimum addend and correct sum squared were significant predictors of RT. As such, this result alone disconfirms the min model for adults, since the model predicts that all of the RT increase should be accounted for by min.

Table 1
Correlation Matrix for Reaction Time (RT) and 10 Predictor Variables: True, Nontie Problems (Experiment 1)

	1	2	3	4	5	6	7	8	9	10
RT	.446	.536	.777	.499	−.692	.737	.775	.042	.137	.064
1		−.111	.577	.577	−.545	.667	.647	.174	−.019	.000
2			.577	.577	−.545	.667	.647	−.019	.174	.000
3				.500	−.742	.866	.899	.100	.100	.067
4					−.674	.866	.781	.100	.100	−.067
5						−.817	−.816	−.089	−.089	.100
6							.971	.116	.116	.000
7								.113	.113	.021
8									−.111	.000
9										.000

Note—1 = first addend, 2 = second addend, 3 = minimum addend, 4 = maximum addend; 5 = number of digits in sum, 6 = correct sum; 7 = sum squared; 8 = odd/evenness of first addend; 9 = odd/evenness of second addend; 10 = odd/evenness of correct sum. The critical values of r for 90 cases are .207 for p < .05 and .270 for p < .01. Odd/evenness was coded .5 for odd, −.5 for even. Number of digits in the sum (Variable 5) was coded .5 for a one-digit sum, −.5 for a two-digit sum.

Table 2
Statistical Summaries of Regression Analyses (Experiment 1)

Step	Variable	Partial F	Prediction Equation	SE	Mult R	F_{reg}	df
			True, Nontie				
1	Min	133.96	RT = 845 + 72.9 (Min)	132.2	.777	133.96	1,88
2	SumSq	7.88, 7.24	RT = 823 + 39.2 (Min) + 1.17 (SumSq)	127.8	.796	75.35	2,87
			True, Nontie, Sum Squared Forced				
1	SumSq	132.45	RT = 822 + 2.26 (SumSq)	132.7	.775	132.45	1,88
2	Min	7.24, 7.88	RT = 823 + 1.17 (SumSq) + 39.2 (Min)	127.8	.796	75.35	2,87
			False, Nontie				
1	Min	21.92	RT = 996 + 36.5 (Min)	163.8	.446	21.92	1,88
2	Split	53.94, 53.72	RT = 1,125 + 46.4 (Min) − 23.4 (Split)	129.5	.711	44.39	2,87
3	SSum	68.45, 56.64, 9.92	RT = 1,046 + 51.7 (Min) − 22.9 (Split) + 7.3 (SSum)	123.4	.746	35.93	3,86

Note—Min = minimum addend; SumSq = sum squared; SSum = stated sum. The final reported step in each analysis represents the best regression solution to that subanalysis. All included variables had significant Fs-to-enter (the last partial F at each step) and were all individually significant in the best equations (p values of .05 or less). The partial Fs are listed for each step and apply to the variables in their order of inclusion.

Two empirical arguments can be presented against Groen and Parkman's (1972) direct-access/counting model. The first of these relates to the split effects, both in the analysis of variance and the regression analysis. Stated simply, these results indicate that part of verification performance involves a decision-like stage of processing, a stage that operates in an equality-inequality judgment fashion. No such stage is included in Groen and Parkman's model, however; Parkman and Groen (1971) do discuss a "comparison" stage, but it was apparently thought to be influenced only by the true-false variable. While the direct-access (or min) model might be "patched" by adding a split-driven decision stage, this modification would still not explain how the false Split 13 problems could be 200 msec faster than the corresponding trues. Since the direct-access model was in fact developed to account for verification performance, this deficiency represents a serious shortcoming.

The second criticism of the direct-access/counting model pertains only to true problems and, as such, does not depend on a decision-stage argument. Recall that the obtained slope of Groen and Parkman's (1972) regression line across min was 20 msec for adults but 400 msec for first-graders. Given these values, and the direct-access assumption of constant retrieval time, the probability of retrieval failure was estimated to be .05. In other words, the fastest 95% of a subject's trials should reveal a min slope of zero, due to direct access; the remaining 5% of the trials should show much slower processing, that is, a 400-msec slope across min.[2]

The present data were examined for exactly this pattern. That is, for each subject individually, the RT at the 95th percentile was determined, and all of that subject's RTs at and above this cutoff point were eliminated. The remaining RTs (91 RTs per subject, on the average) were averaged as before, and resubjected to a regression analysis. This procedure should, according to the direct-access model, result in a nonincreasing RT pattern across problem size, since the analyzed RTs should reflect

only direct-access retrieval. Such was not the case. In fact, in this reanalysis, all of the problem-size variables had significant positive correlations with RT (min = .658, correct sum = .690, sum squared = .720). The best regression solution involved only correct sum squared, $[F_{reg}(1,98) = 105.5, p < .001, R^2 = .52]$. This result is all the more compelling because the test was performed after the extreme scores were removed, and because the cutoff criterion falls somewhat more heavily on large problems, those originally thought to be most responsible for the squared factor's significance.

A more lenient reanalysis was also performed. The slope of RT across min for true problems was 72.9 msec here, not 20 msec as in Groen and Parkman's (1972) report (note, however, that the results here also included sum squared in the best prediction equation). A new, internally consistent percentile criterion was determined with this value (i.e., 72.9 msec/400 msec = .182). With a probability of retrieval failure equal to .182, approximately 82% of the trials should have been direct-access trials. Following the same procedure as above, the 80th (for convenience) percentile RT was used as a cutoff, new means were calculated (based on 76 RTs per subject, on the average), and a new regression was performed. Again, all three of the problem-size variables had significant positive correlations to RT (min = .459, correct sum = .454, sum squared = .447). In fact, a retrieval failure probability of about .50 must be adopted in order to eliminate significant RT increases on the faster set of trials. This is a strikingly unreasonable value to accept in order to support the notion of direct-access retrieval.

The point with these reanalyses is not just that elimination of some proportion of the slowest trials yields similar results to the original analysis, although this is in fact what is obtained (along with a drop in the standard error of the estimate). The point instead is that significant increases in RT are obtained on those trials for which the direct-access model predicts no increases whatsoever. Thus the present results disconfirm the

direct-access/counting model from two standpoints, the lack of an adequate decision stage to explain the split effects and the lack of evidence that direct-access retrieval, at least as defined by Groen and Parkman's (1972) parameters, takes place at all.

While the present results do lend support to the Ashcraft and Battaglia (1978) network retrieval model, they also indicate the need for further specification of the proposed four-stage process model described above. At issue here is the question of serial vs. parallel operation of retrieval and decision. Recall that the Split 13 results showed that for large, and hence slow, problems, an extreme value of split led to faster (by 200 msec) performance than on the same problems presented as true (e.g., 9 + 7 = 3 vs. 9 + 7 = 16). Under the standard assumptions of sequential and independent stages in the additive-factors method (e.g., Sternberg, 1969), retrieval of a sum must be completely finished before any decision-like operation begins. Thus, RTs to both 9 + 7 = 3 and 9 + 7 = 16 must include the same retrieval time, and any remaining difference in RT can be attributed to decision-stage time. Such an interpretation forces the rather problematic conclusion that upon occasion a false/mismatch decision can be made more quickly than a true/match decision. In other words, any sequential explanation of the present results must assert that a false decision based on a large mismatch follows completed retrieval of the sum and is the fastest decision type, whereas true/match decisions, also following completed retrieval, are significantly slower. Slowest of all would be false decisions with lesser degrees of mismatch/split.

We do not favor this explanation, partly because it seems counterintuitive that a decision based on a match would be slower than a mismatch. More critically, this explanation does not suggest why large-mismatch decisions for small problems were not faster than corresponding trues, as these decisions were for large problems. Instead, we favor an interpretation that involves the operation of an evaluation or decision-like process in parallel with retrieval.

Consider the standard four-stage processing model of encoding, retrieval, decision, and response execution that has been presented. Normally, retrieval through the mental representation is required before a decision can be reached, and as such, the retrieval phase would contribute variable amounts of time, depending on the size of the problem. At the end of retrieval, the decision stage performs a comparison between the retrieved sum and the sum stated in the stimulus, yielding a true decision immediately, but a false decision that varies in duration as a function of split. Now add to this model an executive or evaluation process, a process that co-occurs with retrieval. Such a process might perform what we have called global evaluations of the sum of the problem, simply coming up with a progressively more refined region within which the sum is likely to fall. At some point during retrieval, this global evaluation process will

react to an extreme mismatch between the stated sum and the estimated sum, triggering an early decision prior to the completion of retrieval. In Ashcraft and Battaglia's (1978) words, this is a "short-circuiting" of retrieval, based on the magnitude estimations. A simpler possibility here is that subjects responded quickly since the stated sums for large Split 13 problems contained only one digit, whereas the addends were fairly large. In this situation, subjects may have terminated ongoing retrieval based on general arithmetic knowledge that a one-digit sum is highly unlikely with large addends. Even this simpler process, note, involves parallel operation of the evaluation mechanism.

In either case, the decision stage in this model is reserved for the comparison of retrieved and stated sums; that is, the decision stage operates after retrieval is completed. The global evaluation process, however, may operate during retrieval to terminate ongoing processes in situations in which various "rules of arithmetic" are violated. Such an early evaluation process is in principle quite similar to the "fast-no" decision mechanism in the Smith, Rips, and Shoben (1974) model of semantic memory and to the initial search for commonality found in the Collins and Loftus (1975) model. We are proposing simply that other information is retrieved by the subject aside from the correct answer to the problem. This other information, drawn from general knowledge about arithmetic, becomes available during the processing of a stimulus and may influence performance on even simple addition tasks.

From one perspective, this issue of serial vs. parallel processing is of secondary importance here. That is, mental addition generally does not involve the sort of verification of a stated sum that was tested here, and which led to the question of serial vs. parallel processing. For the topic of mental addition, it is undoubtedly more important that the retrieval stage be understood, and the operation of this stage has been addressed specifically by this research. On the other hand, the serial vs. parallel issue is important to a broader understanding of human cognition. While the present research does not provide a definitive answer to this admittedly complex question, the task under investigation does seem potentially useful for this purpose.

EXPERIMENT 2

The purpose of Experiment 2 was to extend the empirical evaluation of mental addition beyond the simple "basic facts," in order to determine the nature of processing for larger addition problems. In order to provide continuity between the studies, Experiment 2 used both simple and complex problems, that is both the basic 100 facts and problems with at least one two-digit addend.

A simple extension of the hypothesized retrieval model to complex addition would claim that RT should be composed of at least the same stages that are neces-

sary for simpler problems, plus other possible mechanisms related to place-keeping and carrying operations. More critically, similar variables should account for RT in both the simple and complex additions. Note that the retrieval model does not suggest that all possible addition problems are stored, but merely that the basic 100 problems are. As such, the extension of the model to complex addition would predict that adults calculate the sum of 14 + 12, for example, by retrieving the separate sums for the components of the problem, 4 + 2 and 1 + 1. In other words, while RT was expected to increase with problem size, it was not expected that the sum of a problem would necessarily provide the best prediction of RT. Instead, the sum of the component addition, or its square, and possibly a constant amount of time for the carry and the 10's column addition might be better predictors of RT.

There is, however, an alternate model for complex addition that deserves attention. Restle (1970) proposed that large complex additions are processed in an analog format, with addends transformed into line segments of varying lengths. His evidence for such a proposal came from an experiment in which subjects had to choose the larger of two quantities, either the sum of two numbers or a comparison number (A + B vs. C). In his results, problem size (essentially the sum of A + B) and split (difference between C and the sum of A + B) exhibited their usual effects. A third factor in the experiment, here termed "addend split" (the difference between A and B) was unique in its effect. With the exception of ties, subjects were faster at adding a relatively small and a relatively large number together, say 72 + 8, than at adding two more nearly equal numbers, 48 + 32. It was primarily this inverse effect of addend split that led to Restle's proposal of the analog process for addition: A shorter line segment is more easily "transported" and "concatenated" to the longer line segment, thereby allowing arrival at the sum more quickly than when a longer line segment is involved.

While the present experiment will evaluate the possibility of this addend split effect (but with somewhat smaller problems than those of Restle, 1970), it should be noted that there are aspects of Restle's results that make a test of his model rather difficult. First, Restle's task seems to involve more steps or processes than the present verification task, in that both a multidigit addition and then an inequality judgment were necessarily part of subjects' RTs. As such, the RT effects did not reflect addition processes alone. Second, the nature of his task and stimuli may have increased the likelihood of responding based on, if not informed guessing, then at least less-than-completed addition processing (see Groen & Parkman, 1972, for a similar argument). It is unclear how such a possibility may have altered the results. In the present paradigm, RTs to true problems are assumed to provide a more direct examination of addition processing per se, as they neither include an inequality judgment nor lead to responding based

primarily on magnitude estimations. Finally, Restle merely graphed the RT results and did not present any statistical analyses. This is a problem for two reasons; one, the effect of addend split was the major evidence of analog processing, but appeared to be a fairly weak RT effect, and two, no information regarding possible interactions was presented. The latter is the more severe problem here, in that Experiment 1 showed a significant interaction of problem size and split. It is not at all clear how the interaction of these two factors, both included in Restle's design, can be interpreted within the framework of his data and model.

Method

Twenty different undergraduates from the same population used in Experiment 1 served as subjects in the 45-min sessions. All were given extra credit for their participation.

The range of correct sums considered in this experiment was 0 to 30. One hundred true problems were randomly selected from the possible pool such that approximately equal numbers came from the three ranges 1-10, 11-20, and 21-30. As a consequence, 46 of the sampled problems were basic 100 facts, and the remaining 54 contained at least one two-digit addend. No attempt was made to eliminate problems requiring the carry operation. Each of the 100 problems was also randomly assigned an incorrect sum that differed from correct by ±1,5 ,9, or 13. The only restrictions on incorrect sums were first that they fall in the allowable range, and second that they be approximately equal in frequency of occurrence to sums in the true condition. Notice that these restrictions did not bias the assignment of problems to the Split 13 condition, as happened in Experiment 1. Aside from these aspects of the stimulus pool, all other procedures were identical to those of the first experiment.

Results and Discussion

Error rate was 2.8%, and extreme scores (Dixon's test) accounted for another 1.3% of the data, a total missing observation rate of 4.1%. Errors and extreme scores were positively related to RT and to problem size and were especially likely in problems with an addend of nine.

The analysis of variance evaluated RT in a problem size (sums of 1-10 vs. 11-20 vs. 21-30) by true-false by split (±1, 5, 9, 13) within-subjects design. All main effects and interactions were significant at the .05 level or beyond and essentially duplicated the patterns obtained in Experiment 1.

Most relevant to present purposes, both the True-False by Problem Size interaction [$F(2,38) = 20.13$, MSe = 16,443.4] and the True-False by Problem Size by Split interaction [$F(6,114) = 3.17$, MSe = 10,074.4] were significant. In the two-way interaction, RT increased sharply, from 959 to 1,378 msec, across increasing problem size in the true problems, but not as sharply for false problems, from 1,093 to 1,332 msec. While this pattern of differences is readily apparent in the three-factor interaction as well (see Figure 2), the qualification of the pattern by split is that the increase in false RT across problem size is progressively attenuated as split increases. In other words, only at the larger values of split did RT to large problems show a speed advantage for false over true stimuli. As the figure shows, true
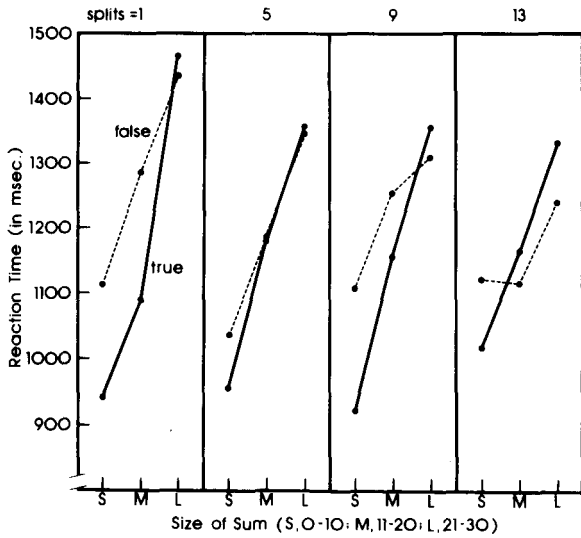
Figure 2. Mean reaction time (RT) to true and false problems as a function of size of sum (S = small, 0-10; M = medium 11-20; L = large, 21-30) and split. (Note—Neither the medium nor the large levels are directly comparable to conditions in Experiment 1.

problems revealed a more or less constant effect of problem size, an approximately 400-msec effect, whereas RT to false problems was simultaneously increased by problem size but decreased by larger and larger splits.

The fact that RT increased dramatically with this larger range of problem size is clearly consistent with previous research but, of course, does not identify the processing responsible for the increase. Accordingly, mean RTs were calculated for each problem and subjected to multiple-regression analysis. In addition to the variables evaluated in Experiment 1, six new variables were coded: the number of digits (one or two) in the first addend and in the second addend, whether or not the problem included a carry operation, the addend split

factor from Restle's (1970) model, and two "component" variables for the complex (nonbasic 100) problems. "Component sum" was the sum of the digits in the one's column of the problem; for 14 + 12, component sum was 6. "Component time" was the mean RT from Experiment 1 to the component addition; using the same example, component time was the mean RT from Experiment 1 to the problem 4 + 2 = 6.

The major regression analysis evaluated true, nontie problems, including both the basic 100 and the larger complex problems (see Table 3 for the correlation matrix). In this analysis, correct sum was selected first, followed by component sum and then the variable that indicated a carry was or was not required (see Table 4 for summaries of the regression results). The significant three-variable equation yielded an intercept estimate of 837 msec, slopes of 17.5 across correct sum and 13.8 across component sum, and an increment of 131 msec when the carry operation was required. The standard error of the equation was 132 msec, and multiple R was .812. Qualitatively similar results, involving component time and correct sum, were found when only the complex true problems were analyzed ("carry" was selected next, but it missed the .05 level of significance).

Given our intent of testing the several verification models of addition, three versions of a general min-like model were also evaluated for these true nontie problems. First, when the standard minimum addend factor was forced into the regression equation alone, it accounted for 43.5% of the variance (as opposed to 54.2% for correct sum alone and 49% for sum squared alone). When minimum addend and the carry factor are forced into the equation, percent variance increases to 57; the significance of minimum addend drops dramatically (partial $F = 63.99$, then 5.42), however, when correct sum is freely selected on the following step (F-to-enter = 17.12). A much more adequate solution

Table 3
Correlation Matrix for Reaction Time (RT) and 14 Predictor Variables: True, Nontie Problems (Experiment 2)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
RT	.513	.519	.660	.630	.258	−.423	−.341	−.665	.736	.702	.453	.501	.622	.479
1		−.018	.558	.592	.286	−.824	−.049	−.529	.665	.635	.170	.274	.584	.084
2			.543	.700	.419	−.046	−.819	−.544	.735	.729	.138	.181	.578	.163
3				.491	−.137	−.591	−.552	−.649	.784	.753	.260	.217	.625	.233
4					.796	−.472	−.569	−.678	.925	.911	.148	.319	.781	.109
5						−.126	−.264	−.320	.507	.513	−.013	.212	.454	−.038
6							.077	.457	−.593	−.580	.144	−.155	−.643	.164
7								.446	−.645	−.650	.158	−.093	−.603	.116
8									−.765	−.654	−.324	−.235	−.673	−.325
9										.976	.218	.321	.828	.179
10											.137	.294	.787	.102
11												.466	−.006	.837
12													.473	.465
13														.043

Note—1 = first addend, 2 = second addend, 3 = minimum addend, 4 = maximum addend, 5 = addend split, 6 = number of digits in first addend, 7 = number of digits in second addend; 8 = number of digits in sum, 9 = correct sum, 10 = sum squared, 11 = component sum; 12 = carry; 13 = component RT; 14 = minimum addend in ones column. The critical values of r for 90 cases are .207 for p < .05 and .270 for p < .01. Number of digits (Variables 6, 7, and 8) were coded .5 for one digit, −.5 for two digits. Carry (Variable 12) was coded .5 if a carry was required, −.5 otherwise. Odd/evenness variables have been omitted here.

Table 4
Statistical Summaries of Regression Analyses (Experiment 2)

Step	Variable	Partial F	Prediction Equation	SE	Mult R	F_{reg}	df
			True, Nontie				
1	CSum	102.87	RT = 849 + 20.7 (CSum)	151.3	.736	102.87	1,87
2	ComSum	99.49, 20.99	RT = 727 + 18.8 (CSum) + 18.8 (ComSum)	136.5	.795	73.75	2,86
3	C	86.58, 9.90, 7.00	RT = 837 + 17.5 (CSum) + 13.8 (ComSum) + 131.4 (C)	131.9	.812	54.93	3,85
			True, Nontie, with Min-Ones				
1	CSum	102.87	RT = 849 + 20.7 (CSum)	151.3	.736	102.87	1,87
2	Min-Ones	112.77, 32.24	RT = 790 + 18.9 (CSum) + 38.6 (Min-Ones)	129.8	.816	86.02	2,86
3	C	96.76, 17.89, 5.25	RT = 866 + 17.8 (CSum) + 31.2 (Min-Ones) + 109.8 (C)	126.7	.828	61.94	3,85
			True, Nontie, Non-"Basic 100"				
1	ComRT	25.01	RT = 535 + .76 (ComRT)	165.2	.589	25.01	1,47
2	CSum	27.33, 19.26	RT = 203 + .68 (ComRT) + 18.9 (CSum)	140.2	.735	27.00	2,46
			False, Nontie, "Basic 100"				
1	CSum	18.53	RT = 952 + 18.1 (CSum)	100.5	.572	18.53	1,38
2	Split	31.50, 9.66	RT = 969 + 23.2 (CSum) − 11.1 (Split)	90.7	.683	16.20	2,37
			False, Nontie				
1	CSum	64.67	RT = 1,001 + 13.0 (CSum)	120.3	.653	64.67	1,87
2	Split	85.29, 12.31	RT = 1,034 + 15.5 (CSum) − 10.2 (Split)	113.0	.706	42.69	2,86
3	Min-Ones	87.60, 16.47, 15.75	RT = 1,003 + 14.6 (CSum) − 10.9 (Split) + 21.7 (Min-Ones)	104.0	.759	38.59	3,85
4	ComRT	11.59, 16.50, 20.64, 7.27	RT = 1,029 + 8.88 (CSum) − 10.5 (Split) + 24.5(Min-Ones) + .10 (ComRT)	100.9	.781	32.89	4,84

Note–CSum = correct sum; ComSum = component sum; C = carry; Min-Ones = min in ones column; ComRT = component RT. The final reported step in each analysis represents the best regression solution to that subanalysis. All included variables had significant Fs-to-enter (the last partial F at each step) and were all individually significant in the best equations (p values of .05 or less). The partial Fs are listed for each step and apply to the variables in their order of inclusion.

is achieved when a new minimum addend variable is used, the smaller of the two addends in the one's column of the problem (for 12 + 9, this new min-ones factor = 2). This min-ones factor joins correct sum and the carry factor in a free stepwise solution, replacing component sum in the process, and improves the prediction equation slightly (see Table 4). In this model, however, correct sum plays a more important role than the min-ones factor (see the partial Fs in the table). This relationship is in close agreement with the equivalent equation in Experiment 1; that is, minimum addend must be augmented by some other problem-size factor like sum to obtain an optimum solution.

These results were judged to be quite consistent with the extended network retrieval model. Our hypothesis was that processing larger additions would involve retrieval of the one's column sum, plus a possible carry operation, and still reveal the general problem-size effect. In fact, just these variables were selected for the best regression equation. As in simpler problems, RT experiences a significant increase as the problem becomes larger. This increase for large problems is affected by both absolute problem size and the size of the component addition. Note here that correct sum and component sum (or correct sum and min-ones) not only contributed independently to the equation but also were in fact minimally interrelated factors (intercorrelations = .218 and .179, respectively). Finally, the significance of the carry variable is particularly interesting, first because it is such an intuitively appealing result, and second

because it emerged so strongly in competition with other structural variables (and despite the fact that only 12 true problems involved the carry at all).

The results from the analyses of ties were consistent with this pattern. Briefly, RT to ties was nearly invariant across sum but was approximately 200 msec slower for large vs. small ties. In other words, processing of true large ties seems to involve retrieval of the component sum, just as for nonties. Note in this case, however, that retrieval of the component sum is a retrieval of a tie and, therefore, contributes a relatively invariant amount of time to the sequence.

Perhaps not surprisingly, the results of the regression analyses on false stimuli were somewhat less clear-cut. This was apparently entirely attributable to the inclusion of the larger, nonbasic fact problems. That is, analysis of just the "basic 100" stimuli yielded a straightforward solution that was quite similar to the Experiment 1 results, involving correct sum and split. When all of the false nontie stimuli are analyzed, however, two more variables enter significantly into the equation, min-ones and component time. Despite the respectable significance of the equation (61% of the variance accounted for), the values of the coefficients do not correspond well to earlier estimates. Of course, this analysis, as well as the analysis of variance, revealed a generally increasing RT pattern across problem size and an attenuation of RT due to split on these larger problems. Apparently, processing of these larger problems is heterogeneous enough within these two effects that no

particularly compelling regression solution is to be expected. As a justification for this, consider the variety of falsification strategies available for a false problem like $14 + 12 = 17$: for example, noting that the stated sum 17 is too close to the addends to be reasonable, retrieving the component sum and noting the mismatch of 6 and 7, or even adding the 1s in the 10s column and noting the mismatch of 2 and 1. Depending on the specific problem, any one of these or other strategies might be interchangeably applied, resulting in RTs that are not clearly indicative of any single process. Further careful manipulation of the various dimensions possible in such problems will be necessary to identify particular strategies.

One final note concerning the results is necessary, pertaining to Restle's (1970) notion about addition of large numbers. Although this experiment did not specifically manipulate addend split, a rather large range was in fact sampled (from equal addend split in the ties up to an addend split of 20). Given this, it is probably important that the addend split factor never approached significance in any of the regression analyses (the highest obtained correlation was .30). This is not to deny that such a factor may play a role in more complex additions; it certainly seems easier to add $72 + 8$ than $48 + 32$, for example. This is merely to suggest that Restle's addend split effect may have been composed of simpler processes. In the above examples, identical additions in the 1s column are followed by very different 10s column additions, 7 + the carried 1 in the first, but 4 + 3 + the carried 1 in the second. In any event, addend split was certainly not found to be the major component of all mental additions predicted by Restle. Its influence, instead, may be limited to larger multidigit additions or may be due to simpler component addition processes.

GENERAL DISCUSSION

We conclude from theoretical and empirical evidence that simple mental addition is largely a memory retrieval phenomenon. Despite children's apparent reliance on counting, adults seem to rely on a stored systematic structure of knowledge for their performance. Explanation of mental addition from the present network viewpoint has several advantages; common experimental results are conveniently described by the structure, performance on more complex problems is predictable, and knowledge and performance under other arithmetic operations can be conceptualized in compatible terms. When investigated under the conditions of a verification task, this structure is revealed most directly in performance on true problems. Performance on false problems in this situation also reveals the operation of decision processes and, possibly, an evaluation stage that co-occurs with retrieval.

The network retrieval/decision model we are proposing here is more than simply a conceptual framework for thinking about mental arithmetic, however. By borrow-

ing some standard assumptions from network approaches in semantic memory, the present model generates some specific predictions about RT performance. One hallmark of network models in semantic memory, for example, is the notion of interrelatedness among concepts (e.g., Collins & Loftus, 1975). In such models, structural connections among concepts provide an explanation for priming effects, either facilitation or inhibition of some semantic judgment due to activation of relevant information (e.g., Ashcraft, 1976; Loftus & Loftus, 1974; Rosch, 1975). Given the assumption of spreading activation, our model would predict several priming/repetition effects, both within and across trials. For example, the model claims that the search for a sum results in activation of related column and row information in the mental table. As such, we would expect speeded verification (or production) of $5 + 6 = 11$ if it were preceded by $7 + 4 = 11$, repetition/priming of a sum, and speeded verification of $5 + 6 = 11$ if it were preceded by $2 + 6 = 8$, repetition/priming of an addend. In both cases, support for the predictions would involve a reduction in the slope across problem size for primed problems; a simple intercept difference might be due to perceptual encoding facilitation (see Ashcraft & Battaglia, 1978). Going one step further, spreading activation also implies decay of that activation, yielding predictions about priming as a function of lag between prime and target. These latter predictions become rather difficult to test, unfortunately, since there are only 10 digits with which to form all problems. As an example, at Lag 3 for the prime and target of $4 + 3 = 7$ and $5 + 2 = 7$, only five other digits are left to form the intervening stimuli (0, 1, 6, 8, 9). Shorter levels of lag, or different decay intervals, should nonetheless support the same conclusions.

Further, the network and priming assumptions predict a slowing of RT when related but competing information is primed. Briefly, since column and row information is presumably activated during the search, using an "activated sum" as the answer to a distractor/false problem should result in an inhibition effect on that distractor. Such an effect was in fact found by Winkelman and Schmidt (1974) when the sum of two digits (7 for $3 + 4$) was used as the false answer in a multiplication fact ($3 \times 4 = 7$). This prediction of intratrial inhibition effects has been tested in a strictly multiplication task (Stazyk, 1980; Ashcraft, Stazyk, & Fierman, Note 2); the prediction was supported, and the results rule out the switches between addition and multiplication as the source of Winkelman and Schmidt's results (see also Duffy & Fisher, Note 3, for a replication of this effect).

One final similarity between semantic and arithmetic research involves the split effect obtained here on false stimuli and its relationship to the semantic distance effect on negatives (Kintsch, 1974). In general, RT to negatives is slowed when two "close" concepts are compared, but it is speeded when the concepts are more distant; for example, RT is slow to "A robin is a sparrow" but faster to a "A robin is a car." In the

present research, we assume that the answer stated in the problem is (usually) compared with the answer retrieved from memory. When the two answers are close, that is, with small values of split or with related information, RT is significantly slowed; when they are more distant (larger splits), RT is speeded. This split effect is, of course, most directly related to the split effect in research on inequality judgments (e.g., Banks et al., 1976; Sekuler, Rubin, & Armstrong, 1971), but it is certainly not limited to just this domain. For example, Holyoak and Walker (1976) found this inverse relationship of RT to "distance" when subjects compared the magnitudes of concepts on the semantic orderings of time, quality, and temperature (see also Friedman, 1978; Kosslyn, Murphy, Bemesderfer, & Feinstein, 1977, for similar results). Walker (1975) has found the same relationship in a verification task; a less "reasonable," that is, a more distant, false sentence yields faster RT. On the assumption that distance in the semantic sense can be roughly equated to distance in the numerical difference sense, the present research seems to fit comfortably within both the mental comparison and the verification approaches to research on mental representations. These similarities also suggest strongly that research on mental arithmetic is of more than casual interest to cognitive psychology. Indeed, it may be that mental arithmetic, with its more controlled acquisition and mastery, and without some of the inconsistencies and fuzziness of semantic knowledge, will prove to be a useful test case for investigating general issues in cognitive representations and their development.

REFERENCE NOTES

1. Ashcraft, M. H., & Fierman, B. *Mental addition in third-, fourth-, and sixth-graders.* Unpublished manuscript, 1980.

2. Ashcraft, M. H., Stazyk, E. H., & Fierman, B. *Network representations of mental arithmetic.* Paper presented at the annual meeting of the Psychonomic Society, Phoenix, Arizona, November 1979.

3. Duffy, S. A., & Fisher, D. L. *The organization and processing of multiplication facts.* Paper presented at the annual meeting of the Midwestern Psychological Association, St. Louis, Missouri, May 1980.

REFERENCES

Ashcraft, M. H. Priming and property dominance effects in semantic memory. *Memory & Cognition,* 1976, 4, 490-500.

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

Banks, W. P., Fujii, M., & Kayra-Stuart, F. Semantic congruity effects in comparative judgments of magnitudes and digits. *Journal of Experimental Psychology: Human Perception and Performance,* 1976, 2, 435-447.

Cohen, J., & Cohen, P. *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, N.J: Erlbaum, 1975.

Collins, A. M., & Loftus, E. F. A spreading-activation theory of semantic processing. *Psychological Review,* 1975, 82, 407-428.

Draper, N. R., & Smith, H. *Applied regression analysis.* New York: Wiley, 1966.

Friedman, A. Memorial comparisons without the "mind's eye." *Journal of Verbal Learning and Verbal Behavior,* 1978, 17, 427-444.

Groen, G. J., & Parkman, J. M. A chronometric analysis of simple addition. *Psychological Review,* 1972, 79, 329-343.

Holyoak, K. J., & Walker, J. H. Subjective magnitude information in semantic orderings. *Journal of Verbal Learning and Verbal Behavior,* 1976, 15, 287-299.

Kintsch, W. *The representation of meaning in memory.* Hillsdale, N.J: Erlbaum, 1974.

Kosslyn, S. M., Murphy, G. L., Bemesderfer, M. E., & Feinstein, K. J. Category and continuum in mental comparisons. *Journal of Experimental Psychology: General,* 1977, 4, 341-375.

Loftus, G. R., & Loftus, E. F. The influence of one memory retrieval on a subsequent memory retrieval. *Memory & Cognition,* 1974, 2, 467-471.

Myers, J. L. *Fundamentals of experimental design* (3rd ed.). Boston: Allyn & Bacon, 1979.

Parkman, J. M. Temporal aspects of simple multiplication and comparison. *Journal of Experimental Psychology,* 1972, 95, 437-444.

Parkman, J. M., & Groen, G. J. Temporal aspects of simple addition and comparison. *Journal of Experimental Psychology,* 1971, 89, 335-342.

Restle, F. Speed of adding and comparing numbers. *Journal of Experimental Psychology,* 1970, 83, 274-278.

Rosch, E. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General,* 1975, 3, 192-233.

Sekuler, R., Rubin, E., & Armstrong, R. Processing numerical information: A choice time analysis. *Journal of Experimental Psychology,* 1971, 90, 75-80.

Smith, E. E., Rips, L. J., & Shoben, E. J. Semantic memory and psychological semantics. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8). New York: Academic Press, 1974.

Stazyk, E. H. *A network approach to mental multiplication.* Unpublished master's thesis, Cleveland State University, 1980.

Sternberg, S. Mental processes revealed by reaction-time experiments. *American Scientist,* 1969, 57, 421-457.

Walker, J. H. Real world variability, reasonableness judgments, and memory representations for concepts. *Journal of Verbal Learning and Verbal Behavior,* 1975, 14, 241-252.

Wike, E. L. *Data analysis.* Chicago: Aldine-Atherton, 1971.

Winkelman, J. H., & Schmidt, J. Associative confusions in mental arithmetic. *Journal of Experimental Psychology,* 1974, 102, 734-736.

NOTES

1. Technically, this tests the null hypothesis that the coefficient (slope) of the new variable equals zero. The F-to-enter test is simply the ratio of additional variance accounted for by the inclusion of the new variable, divided by the still unaccounted for variance. Appropriate formulas are in Cohen and Cohen (1975) and Myers (1979).

2. While both the constant and the multiplicative min term will, of course, have associated variance in their distributions, the absolute size of the multiplicative term, 400 msec, is so large relative to obtained standard deviations/errors of the estimate that no consideration of variance is necessary here.