Involvement of short-term memory
in complex mental calculation

MARIE-PASCALE NOËL, MICHEL DÉSERT, ANNE AUBRUN, and XAVIER SERON
Catholic University of Louvain, Louvain-le-Neuve, Belgium

The aim of this study was to examine the involvement of the short-term memory system in complex mental addition by manipulating the phonological and visual similarity of two numbers to be added. The phonological similarity of the problems appeared to have a major effect on both speed and accuracy. However, the manipulation of visual similarity failed to have any measurable impact. This suggests that the phonological loop, rather than the visual-spatial sketch pad, would be used preferably for temporary storage of addends. An interpretation of these results in terms of the nature of the internal code underlying this task is discussed.

For Hitch (1980), "in arithmetical calculation, the case for supposing some form of short-term memory storage is almost self-evident" (p. 180). Yet, as we will see, things are not so clear and easy. This article reports new results on this topic. The central issue is thus to determine whether doing mental calculation requires the use of the working memory (WM) system (by referring to Baddeley's [1986] model) and, in the affirmative, to identify the component(s) of this system that is (are) recruited.

In dealing with this issue, one must distinguish between simple one-digit and more complex calculations. Indeed, it is generally thought that adults solve single-digit calculations by simply retrieving the solution from long-term memory (e.g., Ashcraft, 1992). In this case, one could consider there to be no need for a mental "working space" and, thus, no WM capacities being used. However, a few studies have examined the involvement of WM in basic arithmetic (Ashcraft, Donley, Halas, & Vakali, 1992; De Rammelaere, Stuyven, & Vandierendonck, 1999; Lemaire, Abdi, & Fayol, 1996). In each study, overloading the central executive had deleterious effects on performance, which led the authors to conclude that there was a necessary involvement of WM resources in these simple calculations. Yet, the introduction of a secondary task taxing the central executive had a main deleterious effect that did not interact with such variables as the complexity of the problem. In other words, this secondary task added a constant amount of time to processing but did not alter the on-going arithmetic performance. Furthermore, the three studies all

used a verification task with simple calculations (e.g., $3 + 5 = 7$?). It is thus possible that the central executive was recruited not so much for the calculation itself (i.e., retrieving the answer from long-term memory) as for the decision process (comparing the retrieved and the proposed answers and selecting one's response and the appropriate response key).

The case of multidigit arithmetical operations is more clear-cut. As early as 1978, Hitch demonstrated that performing such complex addition (e.g., $352 + 259$) involves the WM system mainly for storing the addends (when they are not available on a visual support) and the interim results. Yet, only 16 years later were studies carried out to determine which components of the WM system are used and at which processing stage.

Logie, Gilhooly, and Wynn (1994) compared subjects' performance in adding 2 two-digit numbers under different types of dual-task conditions. When the addends of the problems were presented orally, performance was greatly disrupted by a concurrent random-letter generation task and, more moderately, by concurrent articulatory suppression. Concurrent hand movement or presentation of irrelevant pictures had no effect. The deleterious effect of random generation was replicated when addends were presented visually (for 1 sec), whereas the three other concurrent tasks produced significant but much smaller interference. These results argue for the involvement of the central executive in addition of two-digit numbers. The subvocal rehearsal seems also, but more moderately, implicated (even when addends are presented visually), whereas the visual-spatial component of the working memory, the visual-spatial sketch pad (VSSP), would play a role only when the addends are presented visually. Finally, let us note that the effects of the secondary tasks were not modulated by the task demands (i.e., here, the necessity of a carry procedure).

As was stated above, Logie et al. (1994) concluded that there was a minor involvement of the VSSP, which, furthermore, was restricted to the cases in which the calculations were presented visually (and could thus be re-

stricted to a precalculation stage, such as in the encoding of the visual problem). However, several arguments in the literature seem to indicate the importance of the visual-spatial representations in arithmetical tasks (Hayes, 1973; Pesenti et al., 1998; Seron, Pesenti, Noël, Deloche, & Cornet, 1992; Zago et al., 1999). The aim of the experiments devised by Heathcote (1994) was to test the hypothesis of an involvement of the visual-spatial component of the WM system in calculation. In the first one, three-digit addition problems were compared in two presentation modalities (visual for a short delay or oral), either alone or with three secondary tasks: (1) seeing irrelevant matrices patterns, (2) pressing key pads, and (3) repeating the syllable "the." Furthermore, the answers were to be produced orally, either as a whole or in parts (i.e., announcing the result of each of the subcalculations as soon as they were calculated). All the secondary tasks proved to disrupt the arithmetic performance, with the strongest disruption resulting from articulatory suppression and the least from matrices presentation. However, this effect was modulated by the presence of carries in the problem and the way of responding. Keypress disrupted the arithmetic performance for both types of problems, but more intensely for those requiring carries. The presentation of matrices was disruptive only when the problems required carries, and articulatory suppression was unaffected by the presence of carries. Regarding the way of giving the answers, articulatory suppression, as well as keypress, disrupted the performance for both ways of responding, but more strongly for the partial condition, whereas the presentation of matrices had a similar effect in the two conditions.

According to Heathcote (1994), these results are evidence for the role of the VSSP in the retention of carries, the use of a subvocal rehearsal in the retention of interim results, and the involvement of both the articulatory loop and the VSSP in the retention of the initial problem information. However, these conclusions should be accepted with caution. In particular, the modulation of the articulatory suppression effect according to the way of responding could have a simpler explanation: Producing the partial results orally inevitably stopped the concurrent articulation, which of course, decreased the disruptive effect of saying "the." Thus, the argument for a need for a subvocal rehearsal in the retention of interim results could be put into question. Furthermore, this study suffers from methodological weaknesses (there are only 10 subjects and five items per cell, and correct and incorrect trials are put together for response time [RT] analyses).

In Heathcote's (1994) second experiment, oral or visual addition problems had to be solved under articulatory suppression or with no concurrent task, but this time, the visual similarity of the digits composing the problems was manipulated. As in the previous experiment, articulatory suppression impaired performance. Furthermore, the visual similarity manipulation did not have any effect on RTs and did not interact with any of the other factors (yet only 7 subjects participated in the experiment, and RT analyses were run on both correct and incorrect trials). In the error rates, however, the visual similarity had a small effect, but it appeared only when calculations were presented visually under articulatory suppression. In these conditions, it is not easy to determine whether this effect played a role at the encoding stage or later on, in the resolution process.

In summary, the intervention of the WM system is clear in multidigit calculations. The central executive seems to play a major role (see Logie et al., 1994), yet the deleterious effect of taxing this system does not seem to increase with the difficulty of the calculation. The involvement of a subvocal rehearsal has been suggested by both Logie et al. and Heathcote (1994), but their results could have alternative interpretations. Finally, the role of the VSSP seems to be important in the retention of carries (see Heathcote, 1994). Its involvement in the encoding stage of visually presented problems has also received some support (Heathcote, 1994; Logie et al., 1994), but its role in later processing steps is still open to question. Much still remains to be done in this area, and the aim of the study reported here is to contribute to this issue.

In this work, we were interested in contrasting the participation of the verbal and the visual-spatial slave systems in the realization of the calculation. Therefore, a manipulation of the visual similarity of the addends (such as in Heathcote, 1994) was introduced, as well as a manipulation of the phonological similarity of the addends. It has indeed been shown that the phonological loop is sensitive to this parameter: The retention of phonologically similar material is poorer than the retention of material that presents less phonological similarity (Baddeley, 1986). It must be recalled that Heathcote failed to show any impact of visual similarity. However, the low power of the analyses (very few subjects, results collapsed for both correct and incorrect trials) may be a cause of this failure. Ideally, one would have liked to compare subjects' performance in the four types of stimuli resulting from the combination of high versus low phonological similarity and high versus low visual similarity, but such a selection proved to be very difficult: It was impossible to cross these two criteria while maintaining some control on the difficulty of the problems (in terms of number of carries, size of the addends, and so on). Therefore, two different tasks were set up: a first one, in which phonological similarity was manipulated while controlling visual similarity (Task 1), and another one manipulating visual similarity while controlling phonological similarity (Task 2). Finally, a series of measures of memory spans were taken to examine whether sensitivity to the similarity manipulations was related to the size of the span and, in the affirmative, to identify which memory span (verbal or visual) was the best predictor.

Despite the fact that there are no real theories about the way WM contributes to multidigit addition, one could derive some propositions from the triple-code model proposed by Dehaene (1992; Dehaene & Cohen, 1997). In this perspective, arithmetical facts that have been taught

as verbal routines (basically, the simple additions 1 + 1 to 9 + 9 and the multiplications 1 × 1 to 9 × 9) are stored in that verbal format, and any calculations that have to be made should be translated into that particular format to retrieve the solution. For instance, the multiplication problem 4 × 5 =, presented in the Arabic format, should first be translated into the articulatory–verbal internal code in order to retrieve the corresponding solution. One could thus argue that this transformation of visual Arabic stimuli into a phonological representation is made through the articulatory loop, since this subsystem of the WM is also supposed to be devoted to the translation of visually presented verbal material into a phonological code (Baddeley, Lewis, & Vallar, 1984). By contrast, multidigit calculations are supposed to involve the mental manipulation of a spatial image of the operation in Arabic notation. However, since complex calculations also involve the retrieval of intermediate results, one would predict the involvement of both slave systems while performing complex mental arithmetic: The VSSP would be used as a blackboard on which the multidigit calculations would be performed, and the phonological loop would be used for retrieving the simple arithmetical facts.

EXPERIMENT

The aim of this experiment was to measure the impact of the addends' phonological proximity (Task 1) and visual similarity (Task 2) while adding 2 three-digit numbers. According to Smith (1983), in doing multiplications, exceptional calculators can be described as either auditory (i.e., those who declare "hearing" numbers and multiply them from left to right) or visual (those who declare "seeing" numbers and use cross-multiplication procedures). There is no evidence that such a distinction can also be applied to addition problems or to normal subjects (but see Noël & Seron, 1993). Nevertheless, we thought it was preferable to alleviate as much as possible the factors that could induce either way of dealing with the task. To do this, calculations were not presented in a columnar format (which typically corresponds to the visual method of adding), but rather, addends were presented one after the other. Furthermore, the subjects were explicitly instructed to choose the order in which they wanted to give the answer, either from left to right (e.g., "three hundred and forty-five," which favors a verbal strategy) or from right to left, as in written calculation (e.g., "five, four, three," which favors a visual strategy).

Measures of four different memory spans were also taken: a classical digit span presented in the oral modality, a digit span presented visually, a digit span presented visually with a concurrent articulatory suppression, and finally, a visual-spatial span for nonverbal material.

Method

Subjects

Thirty students from the Catholic University of Louvain agreed to participate in this experiment for course credits. There were 25 females and 5 males, with a mean age of 20.1 years ($SD = 2.2$). All of them were French speaking.

Procedure

All the subjects started with the visual span task, with and without articulatory suppression, then performed one of the two calculation tasks (Task 1 for half of the subjects, Task 2 for the others), followed by the oral digit span and the visual-spatial span tasks and, finally, by the second addition task (Task 2 for half of the subjects and Task 1 for the others).

Procedure for the calculation tasks. Addition problems were presented in the visual Arabic code on a computer screen. Each trial started with a fixation point's appearing at the center of the screen for 1,000 msec. Then, the screen stayed white for another 500 msec. The first three-digit number appeared on the screen for 1,500 msec. Five hundred milliseconds after its disappearance, it was replaced by the second addend, which also remained on the screen for 1,500 msec. Two measures of RT were recorded: one when the subject started producing his/her answer, and the second one when the full answer had been given. These two measures were taken in order to avoid short RTs when the subjects started to say out loud the answer while still computing it (e.g., "eight hundred . . . thirty . . . seven!"). Both of these measures were taken by the experimenter, who pressed the space bar. This methodology was used, instead of the more classical voice-activated relay, in order to avoid spoiled trials owing to subjects' speaking to themselves while computing the result of the addition problem. Errors were also recorded. A 3-sec delay was maintained between the end of the production of the answer by the subject and the beginning of the next trial.

Addition problems were presented in a random order differing for each subject but respecting the following criteria: (1) There were never four problems belonging to the same set of high- or low-similarity condition; (2) there were never two consecutive problems for which the first number (or the second number) ended with the same digit (e.g., 26*4* + 357 and 78*4* + 158); (3) there were never four consecutive problems with the same number of carries.

Procedure for the memory span tasks. Measures of four types of memory span were taken. For each of them, two different series were presented at the same length, starting with a length of two items. The measurement was stopped when the subject failed at two sequences of the same length. The longest sequence that the subject was able, at least once, to repeat correctly was taken as the subject's span.

The *oral digit span* corresponded to the classical memory span: Lists of 2–9 different digits were presented orally to the subjects at the rate of one number word per second. The subject was asked to repeat them in the correct order.

The *visual digit span* was similar to the oral digit span, but the items were presented visually as Arabic digits, rather than as number words. This presentation was controlled by a computer. Each new trial started with a "ready" signal, which was followed, 500 msec later, by the presentation of the first digit. Each digit remained on the screen for 1 sec. After the presentation of the digits series, the subject was asked to repeat aloud the full sequence.

The *visual digit span with articulatory suppression* was similar to the visual digit span, but the subject had to deal with the task while repeating (at a rate of about three per second) the nonsense syllable "bla" during the whole presentation of the Arabic digit sequence. This concomitant vocalization was used to force the subjects to process the Arabic digits visually.

Finally, measure of a *visual-spatial span* was taken with nonverbal material—that is, matrices (developed by Bruyer & Scailquin, 1999). This task was assisted by a computer: Matrices of 4–24 squares were presented on the screen of the computer. In each matrix, half of the squares turned black (on the white background) for 1 sec, one after the other in a random order. The subjects began by viewing matrices of 4 squares, then 6 squares, and so on. Two different matrices of each size were presented. The subject's task was

to draw (and thus localize correctly), on a little booklet, the order of presentation of the dark squares. The measure stopped when he/she failed at two matrices of the same size. The visual span corresponded to the size of the largest matrix the subject was able to draw.

Stimuli

Lists of 48 addition problems of 2 three-digit numbers were generated for each task. For each of these items, scores for phonological and visual similarity were computed.

The phonological similarity was calculated according to the index of phonological similarity proposed by Lhermitte and Desrouesné (1974). This measure was calculated on the corresponding verbal number and corresponded to the following ratio: (number of identical phonemes within the pair of numbers) / (total number of phonemes in the number pair). But for phonemes that were present in the two numbers, the numerator was augmented by two units if it held the same position in the two numbers or by one unit only if it did not. A score close to 0 denoted a low phonological similarity between the two numbers, and a score of 2 indicated a perfect similarity.[1] Since the subjects were French speakers, the verbal numerals were, of course, considered in that language.

The index of visual similarity was based on visual similarity data reported by Campbell and Clark (1988). These authors asked subjects to rate the visual similarity between two digits on a 7-point scale going from 1 (*weak similarity*) to 7 (*high similarity*). The matrix they obtained was not perfectly symmetrical (e.g., the similarity between 1 and 5 [= 1.13] was different than the one between 5 and 1 [= 1.25]), but since judgments on symmetrical pairs were highly correlated (.96), we decided to consider, for each pair of digits, the mean between the scores for the two symmetrical pairs (i.e., considering 1.19 for both 1–5 and 5–1). Furthermore, a score of 7 (very high similarity) was given for each pair of identical digits. For each of the calculations used in the task, we calculated the mean of the visual similarities between each digit and the five others in the addition problems.[2]

Stimuli for the addition problems in Task 1. The 48 addition problems involved 2 three-digit numbers, with no repetition of the same digit (within, as well as between, the numbers) and no zero (in order to maintain a high complexity). These items were distributed in two categories: Problems with a low phonological similarity had a mean index of phonological similarity of .37 (SD = .07), and those of high similarity had a score of .88 [SD = .08; $t(46) = 6.46$, $p < .0001$]. The two sets of items did not differ in terms of number of words (M = 8 in the two sets), number of syllables (M = 10 for both sets), and visual similarity [M = 3.29, SD = .37, and M = 3.12, SD = .37, for the low and high phonological similarity sets, respectively; $t(46) = 0.99$, n.s.]. The measures of visual and phonological similarity did not correlate with one another [$r(46) = -.09$, n.s.].

Each set included 12 addition problems requiring a carry on both the unit and the decade and 12 items with no carry on either the unit or the decade.

Stimuli for the addition problems in Task 2. Forty-eight addition problems of 2 three-digit numbers were generated. The numbers could not involve a zero and had to present at least five different digits. All the sums were between 500 and 1,000. Two sets of items were distinguished according to their high versus low visual similarity. The 24 addition problems of the low visual similarity set presented a mean index of visual similarity of 2.70 (SD = 0.16), whereas a mean of 4.30 (SD = 0.22) was obtained for the high visual similarity set [$t(46) = 12.66$, $p < .0001$]. These item sets did not differ in terms of phonological similarity [M = 0.68, SD = 0.23, and M = 0.62, SD = 0.14, respectively; $t(46) = 0.53$, n.s.], and the visual and phonological proximity measures did not correlate with one another [$r(46) = -.14$, n.s.].

Among each of these sets, there were 12 problems requiring two carries (both on the units and the decades) and 12 for which only one carry was necessary (six times on the units and six times on the decades).

Results

Addition

Despite the fact that the subjects could choose to express their answers from left to right (starting with the thousands and moving on to the hundreds, tens, and units) or right to left (starting with the units), they all responded the same way—that is, by producing a verbal answer corresponding to the left-to-right order (i.e., the usual order for verbal numerals). However, two observations are worth making. First, one of the subjects (A5) was muttering while doing the calculation, and the steps of her calculation clearly revealed that she was calculating from right to left (starting with the addition of the units, then moving on to the tens and, finally, the hundreds). However, she nevertheless spoke out the answer in the left-to-right order. The case of B14 is also interesting. In the debriefing, this person reported that she started the task by calculating from right to left, as in written calculations, but then changed her strategy because she was forgetting the digits while calculating.

Response times analysis. Since the two measures of speed correlated highly with one another[3] and gave rise to the same profile of results, only the analysis of the global time necessary to produce the full answer will be reported here.

Addition, Task 1. Median RTs were calculated for each subject on correct responses on each of the four cells (i.e., two levels of phonological similarity and two levels of carry). The data were entered in a repeated measures analysis of variance (ANOVA) with two factors: phonological proximity (low vs. high) and number of carries (0 vs. 2). Addition problems with a high phonological similarity were processed more slowly than those with a low similarity [$F(1,29) = 35.496$, $p < .0001$], but this effect arose only for those items that required two carries. Indeed, the main effect of carry did not reach statistical significance, but the interaction was significant at $p = .0001$ [$F(1,29) = 20.079$]. There was no difference between the low- and high-similarity sets for the no-carry condition [$t(29) < 1$], but this difference was highly significant for the two-carry condition [$t(29) = 27.323$, $p < .00001$; see Table 1].

Addition, Task 2. Median RTs for correct responses were entered in a 2 × 2 repeated measures ANOVA, with visual similarity and carry as factors. Only the carry had a significant influence on RTs [$F(1,29) = 10.081$, $p = .0035$], indicating faster RTs for addition problems with one carry than for those with two carries (10,819 and 12,585 msec, respectively). There was no effect of visual similarity ($F = 2.09$, $p > .1$) and no interaction between the two factors ($F = 1.72$, $p > .1$).

Global analysis. The aim of the subsequent analyses was to evaluate the relative importance of the complexity of the calculations, of phonological proximity, and of visual proximity in predicting the subjects' addition speed.

Table 1
Mean Number of Errors (with Standard Deviations) and Mean Median Response Times
(RTs, in Milliseconds) as a Function of the Number of Carries and the
Phonological and Visual Proximity in Task 1 and Task 2, Respectively

| | Low Similarity | | | | High Similarity | | | |
| | Errors | | RTs | | Errors | | RTs | |
Number of Carries	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
			Phonological (Task 1)					
0	3.83	2.53	9,425	5,607	4.13	2.33	10,104	7,216
2	4.67	2.22	8,328	6,418	7.20	3.10	13,425	7,251
			Visual (Task 2)					
1	4.43	2.36	11,095	7,556	4.90	2.25	10,543	6,234
2	5.40	2.76	13,407	10,965	5.13	2.65	11,763	6,787

They were conducted on the reassembled data collected in Tasks 1 and 2.

A stepwise regression analysis was calculated, with the median RT (calculated on all the subjects' correct answers) for each item (48 items from Task 1 and 48 from Task 2) as the dependent variable. Six predictors were considered. Three of them characterized the complexity of the addition problems: (1) the solution to the calculation (e.g., 615 for 254 + 361), (2) the sum of all the subadditions involved [e.g., for 254 + 361, we have (2+3) + (5+6) + (4+1) = 21], and (3) the number of carries required for the calculation. Two predictors concerned the verbal representation of the addends: (1) the number of syllables in the two addends and (2) the index of phonological similarity. The last predictor was the index of visual similarity.

Three variables were selected by the analysis [R^2 = .21.; $F(3,92) = 8.22$, $p = .0001$; see Table 2 for the intercorrelation matrix]: Phonological similarity entered first (explaining 10% of the variance; $B = 4,062.8$, $p = .002$), followed by the number of carries (which increased the explained variance by 5%; $B = 1,521.9$, $p = .0007$) and then by the solution of the calculation (which increased the explained variance by another 6%; $B = 2.81$, $p < .007$).[4] When both the number of carries and the solution were partialled out, the phonological similarity still explained 8.6% of the variance.

Error rate analysis. For *addition, Task 1*, errors were very frequent. On average, the subjects erred on 19.8 out of the 48 problems (41.3%). Number of errors was submitted to a repeated measures ANOVA with two factors: phonological proximity (low vs. high) and number of

carries (0 vs. 2). More errors were produced in response to addends with a high rather than a low phonological proximity [$F(1,29) = 17.366$, $MS_e = 3.467$, $p = .0003$]. Problems with carries led to more errors than those without [$F(1,29) = 21.668$, $MS_e = 5.265$, $p = .0001$]. And these two factors interacted with one another [$F(1,29) = 11.622$, $MS_e = 3.219$, $p = .002$], so that the effect of the phonological similarity appeared only when the calculation involved two carries. For the no-carry addition problems, there was no difference between the low- and the high-similarity sets [$t(29) < 1$], but this difference was significant for the addition problems with two carries [$t(29) = -5.291$, $p < .00001$; see Table 1].

Addition, Task 2. The average number of errors was 19.9 (41.3%). A repeated measures ANOVA, with visual proximity (low vs. high) and number of carries (one vs. two) as factors, indicated a main effect of number of carries [$F(1,29) = 4.19$, $MS_e = 2.58$, $p = .05$; mean of 9.33/24 for one-carry and 10.53 for two-carry addition problems] but no effect owing to visual similarity [$F(1,29) < 1$] and no interaction between these two factors [$F(1,29) = 2.52$, $p > .1$].

Global analysis. When the data collected from the two tasks are put together, a mean number of 40 errors per subject ($SD = 14.75$) is obtained. To identify the characteristics of the items that influenced this error rate, a stepwise regression analysis was carried out on the number of errors committed on each item, with the same predictors as those used in RT analysis: (1) the solution to the calculation, (2) the sum of all the subadditions involved, (3) the number of carries required for the calculation, (4) the number of syllables in the two addends,

Table 2
Intercorrelation Matrix Between the Total Number of Errors Produced on an Item
and the Six Predictors of the Regression

	Sum	Solution	Carries	Syllables	Phono	Visual
Total errors	.38	−.03	.42	.13	.49	.13
Median response time	.19	.04	.26	.13	.32	−.05
Sum	1	.23	.42	.35	.10	.63
Solution		1	−.60	.19	−.13	−.10
Carries			1	.10	.14	.31
Syllables				1	−.09	.20
Phono sim					1	−.09
Visual sim						1

(5) the index of phonological similarity, and (6) the index of visual similarity. Three of these variables were selected [$R^2 = .45$; $F(3,92) = 25.063$, $p < .00001$; see Table 2 for the intercorrelation matrix]. The phonological similarity entered first ($B = 8.40$, $p < .00001$) and explained 24% of the variance, then the number of carries (which increased the explained variance by 12%; $B = 2.99$, $p < .00001$), and finally, the solution of the calculation (which added another 9% of explained variance; $B = 0.0045$, $p = .0002$). When partialling out the effects of carries and of the solution, the correlation between the number of errors and the phonological similarity remained highly significant [$r(94) = .504$, $p < .0001$].

Type of errors. Since there has never been any description of the type of errors produced by young adults in three-digit mental addition, we report some analyses of our corpus. Among the 1,200 errors produced in the two tasks, there was 177 nonresponses (14.75%), 72 partial responses (6%; i.e., the subject started giving the response out loud but then stopped in the middle, usually because one part of the response, or of the problem, was forgotten), and 951 calculation errors (79.25%).

Calculation errors were examined more deeply. First, we considered the position of the digit(s) that was (were) incorrect in the answer (e.g., 368 + 457 = *837* instead of 825; there are errors on both the units and the tens positions). In our corpus, errors occurred equally often on the units ($M = 14.41 \pm 9.23$), tens ($M = 16.93 \pm 9.38$), and hundreds ($M = 14.60 \pm 7.19$) digits but much more rarely on the thousand digits ($M = 2.27 \pm 2.63$).[5] Let us remember here that four-digit solutions could only have a 1 in the thousand position.

Second, we examined the numerical distance between the erroneous and the expected digits. For instance, in 146 + 785 = *831* instead of *931*, there was a distance of 1 between 8 and 9, whereas in 652 + 943 = 1,*5*55 instead of 1,*5*95, the distance between 5 and 9 was 4. This numerical distance proved to vary significantly with the position of the erroneous digit within the number. In particular, this distance increased as we moved toward the rightmost position: The distance was 1.63 for the hundreds digit, 2.2 for the tens, and 2.73 for the units [$F(2,58) = 22.08$, $p < .00001$; all the 2 × 2 comparisons were significant at $p < .007$].[6] This progression could support at least two interpretations. First, since our subjects all seemed to calculate from left to right, one could imagine that as the calculation process proceeded, the addends of the problem tended to fade away from memory, thus provoking larger errors. Second, one could imagine that, in parallel with the precise calculation of the sum, the subjects also would have had a global estimation of the addition problem's result. From this perspective, a distance error would be less detrimental if it appeared in the units position, rather than in the tens or the hundreds position.[7]

Finally, we wondered whether errors were frequently due to a malfunctioning of a carry procedure. Error at that level could manifest itself in two ways: carrying a 2 rather than a 1 or failing to do the carry. In both cases,

the digit produced would have a distance of ±1 from the expected one. Consequently, we compared, for each digit position, the ratio between ±1 and more distant errors, assuming that this ratio should be greater when carry was required than when it was not. For the units, no such difference was expected, since no carries were necessary.[8] The pattern of results was mixed. Indeed, for the hundreds, the ratio of ±1 errors was higher for the items that required a carry on the hundreds position ($M = 67.83\%$, $SD = 18.60$) than for those that did not [$M = 50.10\%$, $SD = 27.95$; $t(29) = 2.96$, $p = .006$]. But for the tens, the ratio of ±1 errors was equal for the items that required a carry on the tens ($M = 36.50\%$, $SD = 15.79$) and for those that did not [$M = 38.16\%$, $SD = 25.55$; $t(29) = -.35$].

In summary, calculation errors appeared more often on the units, tens, and hundreds positions than on the thousands position. Second, the numerical distance between the erroneously produced digit and the expected one tended to increase as we moved toward the end of the oral answer. Third, carry procedures in the hundreds position tended to increase the error rate, which was not the case for carry procedures in the tens position.

If we now come back to the main purpose of this research—that is, the impact of a manipulation of the phonological or the visual similarity of the addends on their addition—we know so far that the manipulation of phonological similarity (but not of visual similarity) had an impact on RTs and error rates. We could now raise the question of whether these manipulations had any effect on the errors themselves. As was described above, calculation errors could be more or less distant from the expected answer. This distance can be approached by two indices: (1) the number of digits that are common between the erroneous response produced by the subject and the expected sum (e.g., 146 + 785 = 921 instead of 931, two correct digits, whereas in 146 + 785 = 823, there are none) and (2) the numerical distance between each digit of the given response and the expected one [e.g., in 368 + 457 = 837 instead of 825, the mean distance is $(0+1+2)/3 = 1$].

In order to examine whether the visual and phonological similarity of the addends had some impact on the errors produced, stepwise regression analyses were computed, with these two indices of distance as dependent variables and the following six variables as predictive factors: (1) solution to the calculation, (2) sum of all the subadditions involved, (3) number of carries required for the calculation, (4) number of syllables in the two addends, (5) index of phonological similarity, and (6) index of visual similarity.

For the mean number of digits that are common between the erroneous response produced by the subject and the expected sum, two variables were retained: phonological similarity ($B = -.27$, $p < .004$; $r^2 = .07$) and number of syllables ($B = -.07$, $p < .015$; increases the explained variance by 6%). For the mean distance between each digit of the erroneous response and the corre-

sponding digit in the correct sum, three variables were selected: number of carries ($B = -.25$, $p < .0001$; $r^2 = .16$), sum of the subadditions ($B = .03$, $p < .002$; increases the explained variance by 8%), and phonological similarity ($B = .42$, $p < .004$; increases the explained variance by another 7%).

If these two analyses did not show a great amount of explained variance, they both retained one of the factors we manipulated (i.e., phonological similarity) as a significant predictive factor. Thus, the higher the phonological similarity in an addition problem and the longer it is in terms of number of syllables, the lower the number of shared digits between the produced and the correct sums. Furthermore, when we calculate the distance between each digit of the error and the expected digit in the answer, we note that this distance tends also to increase as a function of the phonological similarity of the addends.

Addition and Memory Spans

Memory spans. Four measures of short-term memory were taken. The visual-spatial span using matrices did not present any significant correlation with any of the three other span tasks (not even with the visual digit span with concurrent articulatory suppression; all Spearman correlations below .1). By contrast, all the measures involving numbers were highly correlated with one another: Oral digit span was correlated with visual digit span ($r = .64$) and with visual digit span under articulatory suppression ($r = .61$); these two visual digit spans also were correlated with one another ($r = .53$). A principal components analysis was run on these three digit spans. The first factor was highly correlated with each of these three spans ($r > .83$ for each of them) and explained 73% of the variance. This factor (hereafter called *score of digit span*) will be used later on, as a unique measure, instead of taking the three measures separately. This score of digit span was unrelated to the visual-spatial span ($r = .0037$).

Memory spans and additions. The relationship between the subjects' short-term memory capacities and their addition performance was investigated by several correlations. First, correlations were calculated between the memory spans and the subjects' median RTs (calculated on the correct answers to the 96 items), but none of the coefficients reached statistical significance. Then, correlations were calculated with the subjects' number of errors: Moderate correlations were found with the score of the digit span [$r(28) = -.31$, $p < .05$] and the visual-spatial span [$r(28) = -.35$, $p < .05$], indicating that the subjects with poor short-term memory capacities tended to produce more addition errors.

More important, we wanted to see whether the sensitivity to our manipulation was associated with poor or good short-term memory performance. To that aim, an index of sensitivity to our manipulation was calculated as follows: (high similarity condition − low similarity condition) × 100 / (high similarity condition + low similarity condition). This formula was applied to both the number of errors and RTs, separately for Task 1 and Task 2 addition problems. The only significant correlation ($p < .05$) was between the sensitivity to phonological similarity (thus, on Task 1 addition) measured on error rates and the score of digit span ($r = -.37$). Thus, the subjects who were the most sensitive to the manipulation of the phonological similarity of the addends (i.e., those who produced more errors in the high-similarity condition, relative to the low-similarity condition) tended to be those who had the lower digit spans.[9] By contrast, the short-term memory performance did not predict the sensitivity to the manipulation of the visual similarity of the problems.

DISCUSSION

A few researchers have tried to determine the role of the WM system in performing calculations. Nearly all of them have examined this question by measuring the impact of a secondary task, whose aim was to overload one of the components of the WM system, on the calculation performance. In this study, our aim was to examine the role of the two slave memory systems in a multidigit addition task by using a different approach—that is, by manipulating the visual or the phonological proximity of the problem's addends. This methodology goes totally unnoticed by the subjects and avoids a usual weakness of the secondary task methodology—that is, the great difficulty of measuring the secondary task's complexity.

The results obtained were clear-cut. Manipulation of the visual similarity of the addends had no effect on the subjects' performance, thus replicating Heathcote's findings. However, the phonological similarity of the addends greatly affected the subjects' calculations. Indeed, increasing the phonological similarity of a complex addition problem tended to lengthen RT, to increase the number of errors, and to give rise to errors that were more distant from the correct sum (in terms of both number of correct digits and numerical distance between the produced digits and the expected ones). Finally, measures of short-term memory capacities indicated that not all the subjects were equally affected by this manipulation of phonological similarity. Indeed, it was those who had the lower digit spans that were more sensitive to the addends' degree of phonological similarity. These results are even more impressive given the fact that addition problems were presented visually in the Arabic code and not orally.

The great impact of the manipulation of phonological similarity adds to the other arguments in the literature indicating that the phonological loop plays a major role in mental addition, whereas the VSSP does not seem to be particularly involved. Thus, in mental computation, the storing space is verbal, and if calculations are presented visually in digits, a phonological recoding process first

takes place. Let us note that this verbal recoding of Arabic digits has also been evidenced in noncalculation tasks. In particular, Brysbaert (1995) showed that it was also the case in a simple memory scanning task (i.e., three Arabic numbers are presented successively and have to be stored in short-term memory by the subject in order to indicate whether a comparison stimulus belongs to the memory set or not).

Our findings do not correspond to the predictions derived from the triple-code model proposed by Dehaene (1992; Dehaene & Cohen, 1997). In this model, multidigit calculations are supposed to require the mental manipulation of a spatial image of the operation in Arabic notation. Consequently, this is an effect of the manipulation of the visual similarity of the addends that is mostly expected but that was not found here. For Dehaene, the phonological loop would be used merely for retrieving the simple arithmetical facts necessary to realize the intermediate calculation steps. Yet, the great impact of the phonological similarity of the addends indicates that the phonological loop also intervenes as a storing space for the problems themselves (this, of course, does not imply that this system would not also play a role in other computing stages, such as, for instance, in the retrieval of the individual arithmetical facts).

Finally, we would like to consider the question of individual differences. Previously, we (Noël & Seron, 1993) have proposed that some individuals might have a privileged code for keeping the information while doing some processing. If our results suggest that the verbal code is used by most people, they do not allow us to exclude the possibility of the visual code's being used by a few subjects. For instance, the VSSP might be preferentially used by subjects who have developed a visual number form (see Seron et al., 1992) or by expert abacus subjects (see Hatano & Osawa, 1983; Hatta, Hirose, Ikeda, & Fukuhara, 1989). If we now consider each individual profile in our study, we can also detect a few people who seem to be sensitive to the visual proximity manipulation. In particular, A15 produced 2 errors in the low visual proximity condition, but 10 in the high visual similarity condition. Similar differences were obtained for A9 (1 and 5 errors respectively) and for A14 (7 vs. 13 errors).[10]

If our results argue for a verbal recoding of visual Arabic digits in mental addition, they may not speak directly to the nature of the code used for storing arithmetical facts in long-term memory. Indeed, for some authors, there is no real difference between the types of information that are stored in short-term versus long-term memory, the main difference being only their degree of activation (e.g., Clark & Campbell, 1991). For others, however, these two types of information may differ in the nature of the code underlying their storage. For instance, one can think of short-term information as being stored under a verbal or visual-spatial format but hypothesize that all information is stored under an abstract or propositional code in long-term memory (e.g.,

McCloskey, Caramazza, & Basili, 1985). But will we ever be able to study representations other than those which are activated?

REFERENCES

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, **44**, 75-106.

Ashcraft, M. H., Donley, R. D., Halas, M. A., & Vakali, M. (1992). Working memory, automaticity, and problem difficulty. In J. I. D. Campbell (Ed.), *The nature and origins of mathematical skills* (pp. 301-329). Amsterdam: Elsevier.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D., Lewis, V. J., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology*, **36**, 233-252.

Bruyer, R., & Scailquin, J.-C. (1999). Assessment of visuo-spatial short-term memory and effect of age. *European Review of Applied Psychology*, **49**, 175-180.

Brysbaert, M. (1995). Arabic number reading: On the nature of the numerical scale and the origin of phonological recoding. *Journal of Experimental Psychology*, **124**, 434-452.

Campbell, J. I. D., & Clark, J. M. (1988). An encoding complex view of cognitive number processing: Comment on McCloskey, Sokol and Goodman (1986). *Journal of Experimental Psychology: General*, **117**, 204-239.

Clark, J. M., & Campbell, J. I. D. (1991). Integrated versus modular theories of number skills and acalculia. *Brain & Cognition*, **17**, 204-239.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, **44**, 1-42.

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex*, **33**, 219-250.

De Rammelaere, S., Stuyven, E., & Vandierendonck, A. (1999). The contribution of working memory resources in the verification of simple mental arithmetic sums. *Psychological Research*, **62**, 72-77.

Hatano, G., & Osawa, K. (1983). Digit memory of grand experts in abacus-derived mental calculation. *Cognition*, **15**, 95-111.

Hatta, T., Hirose, T., Ikeda, K., & Fukuhara, H. (1989). Digit memory of sorobane experts: Evidence of utilization of mental imagery. *Applied Cognitive Psychology*, **7**, 265-281.

Hayes, J. R. (1973). On the function of visual imagery in elementary mathematics. In W. G. Chase (Ed.), *Visual information processing* (pp. 177-214). New York: Academic Press.

Heathcote, D. (1994). The role of visual-spatial working memory in the mental addition of multi-digit addends. *Cahiers de Psychologie Cognitive*, **13**, 207-245.

Hitch, G. J. (1978). The role of short-term working memory in mental arithmetic. *Cognitive Psychology*, **10**, 302-323.

Hitch, G. J. (1980). Developing the concept of working memory. In G. Claxton (Ed.), *Cognitive psychology: New directions*. London: Routledge & Kegan Paul.

Lemaire, P., Abdi, H., & Fayol, M. (1996). The role of working memory resources in simple cognitive arithmetic. *European Journal of Cognitive Psychology*, **8**, 73-103.

Lhermitte, F., & Desrouesné, J. (1974). Paraphasies et jargon-aphasies dans le langage oral avec conservation du langage écrit: Genèse des néologismes [Paraphasia and jargon aphasia in oral language with preserved written language: Origins of neologisms. *Revue Neurologique*, **130**, 21-38.

Logie, R. H., Gilhooly, K. J., & Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Memory & Cognition*, **22**, 395-410.

McCloskey, M., Caramazza, A., & Basili, A. (1985). Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brain & Cognition*, **4**, 171-196.

Noël, M.-P., & Seron, X. (1993). Arabic numeral reading deficit: A

single case study. Or, when 236 is read (2306) and judged superior to 1258. *Cognitive Neuropsychology*, **10**, 317-339.

PESENTI, M., TZOURIO, N., DUROUX, B., SAMSON, D., BEAUDOUIN, V., SERON, X., & MAZOYER, B. (1998). Functional anatomy of mental calculation in a calculating prodigy [Abstract]. *NeuroImage*, **7**, 822.

SERON, X., PESENTI, M., NOËL, M.-P., DELOCHE, G., & CORNET, J.-A. (1992). Images of numbers or "When 98 is upper left and 6 sky blue." *Cognition*, **44**, 159-196.

SMITH, S. B. (1983). *The great mental calculators*. New York: Columbia University Press.

ZAGO, L., PESENTI, M., MELLET, E., BRICOGNE, S., SERON, X., BEAUDOIN, V., LOCHON, P., MAZOYER, B., & MAZOYER, N. (1999). Functional anatomy of simple and complex mental calculation using PET [Abstract]. *NeuroImage*, **6**, 325.

NOTES

1. Let us take an example: $152 + 418$. In French, this corresponds to "cent cinquante-deux + quatre cent dix-huit," or, in phonetic writing: [s - a̱ - s - _ - .k - a̱ - t - d - _] and [k - a - t - r - s - a̱ - d - i - z - u - i - t]. There is a total of 21 phonemes. Some of them are present more than once: There are two replications of the [s], two of the [a̱], one of the [k], two of the [t], one of the [d], and one of the [i]. However, an [a̱] appeared at the sixth position within the two numerals. The score is thus equal to 8 (replications) + 2 (the replication in the same position), divided by 21: $10/21 = .476$.

2. For instance, for $152 + 418$, the average visual similarity is 2.6 between 1 and the other five digits (1–5, 1.19; 1–2, 1.19; 1–4, 2.62; 1–1, 7; 1–8, 1), 2.58 between 5 and the other digits, 2.43 between 2 and the other digits, 2.24 between 4 and the other digits, 2.6 between 1 and the other digits, and 2.34 between 8 and the other digits. This thus gives a mean of 2.46 = index of visual similarity for the item $152 + 418$.

3. The Bravais–Pearson correlations calculated between the two speed measures in each of the four cells of the design varied between .92 and .99.

4. The correlation between visual similarity and the median RTs was not significant ($r = -.05$).

5. The difference between these four positions is significant [$F(3,87) = 45.18, p < .00001$] but is due only to the differences between the rate of errors in the thousand position versus the other positions (paired t test using Bonferroni correction, $p < .00001$).

6. For the digit in the thousands position, errors were rare, and all but one of them were distant at ± 1. This can easily be explained by the fact that none of the addition problems had an answer equal or superior to 2,000. The digit in the thousands position could thus be either absent or equal to 1.

7. However, one should acknowledge that these two interpretations would also predict a similar trend in the error rate—that is, more errors on the units position than on the tens position and so on, which is not precisely what was observed.

8. When the sum of the two units is larger than 10, the carry has to be added on the tens, rather than on the units.

9. Note that no new significant correlations appeared when indices of sensitivity to the manipulation were calculated separately for each number-of-carry condition.

10. It should, however, be noted that all these subjects also show the classic sensitivity to phonological similarity (4 vs. 7 errors for A9, 9 vs. 16 errors for A14, and 5 vs. 7 errors for A15) and that none of these subjects showed any indication of using a visual strategy for calculating (i.e., using a right-to-left order; see the beginning of the Results section).