Retrieval or nonretrieval strategies in mental arithmetic? An operand recognition paradigm

CATHERINE THEVENOT
University of Geneva, Geneva, Switzerland

AND

MURIEL FANGET AND MICHEL FAYOL
Université Blaise Pascal, Clermont-Ferrand, France

According to LeFevre, Sadesky, and Bisanz (1996), averaging solution latencies in order to study individuals' arithmetic strategies can result in misleading conclusions. Therefore, in addition to classical chronometric data, they collected verbal reports and challenged the assumption that adults rely systematically on retrieval of arithmetic facts from memory to solve simple addition problems. However, Kirk and Ashcraft (2001) questioned the validity of such a methodology and concluded that a more appropriate method has to be found. Thus, we developed an operand recognition paradigm that does not rely on verbal reports or on solution latencies. In accordance with LeFevre et al., we show in a first experiment that adults resort to nonretrieval strategies to solve addition problems involving *medium* numbers. However, in a second experiment, we show that high-skilled individuals can solve the same problems using a retrieval strategy. The benefits of our paradigm to the study of arithmetic strategies are discussed.

Numerous researchers have indicated that adults solve simple arithmetic problems more or less exclusively by direct retrieval of the answer from a network of associations stored in long-term memory (Ashcraft, 1992, 1995; Campbell, 1995; for a review, see McCloskey, Harley, & Sokol, 1991). It is indeed widely accepted that performance of young children in arithmetic is based on counting or other procedural strategies and that these procedures are gradually replaced by direct memory retrieval (Ashcraft, 1992; Barrouillet & Fayol, 1998; Campbell & Oliphant, 1992; Lemaire, Barrett, Fayol, & Abdi, 1994; Siegler, 1996; Widaman & Little, 1992; Widaman, Little, Geary, & Cormier, 1992; for a review, see Geary, 1994). This fact was first supported by Groen and Parkman (1972) from latency data in children and adults. The authors showed that for primary school children, response times (RTs) for simple addition problems (e.g., $4 + 3$) increase linearly with the size of the smaller operand. This result was the first to provide evidence for the use of the *min strategy* by children. This strategy consists in counting on from the larger of the two operands by the number indicated by the smaller of the operands (Carr & Jessup, 1995; Siegler, 1987; Siegler & Crowley, 1994). Adults also show a significant increase in response latencies as a function of the size of the operands, but this increase is much smaller than that in children. Moreover, unlike in children, adults' RTs form a curvilinear function that is best explained by the square of the sum or the product of the operands than by

the size of the operands. These differences between adults and children have been interpreted as evidence that, unlike children, adults use fast and efficient retrieval from memory to solve simple addition problems.

However, LeFevre, Sadesky, and Bisanz (1996) stressed that, as has already been mentioned for children (Siegler, 1987, 1989), averaging solution latencies across trials that involve different procedures can result in misleading conclusions about how adults solve problems. These authors note that the researchers who have used the more direct approach of asking participants to report their procedure have challenged the strong assumption that adults always use retrieval for simple addition problems. For example, Svenson (1985) showed that adults were certain that they had used a retrieval strategy on simple addition problems on only 78% of the trials (for similar results, see also Geary, Frensch, & Wiley, 1993; Geary & Wiley, 1991). Then, to address the issue of how the selection of procedures varies across problems and participants, LeFevre et al. collected trial-by-trial reports of procedure, in addition to the classical chronometric data. The authors concluded that the importance of retrieval had been overemphasized in models of adult performance. Indeed, in the case of addition problems, 81% of their participants had used two or more of the counting, retrieval, and decomposition procedures in order to solve simple problems (i.e., both the addend and the augend inferior to 10). In fact, retrieval was the most frequently used procedure for problems with sums lower

C. Thevenot, catherine.thevenot@pse.unige.ch

than 10 (about 83% of use), but the use of transformation procedures increased dramatically for problems with sums greater than 10 (about 46%, if we do not consider the ties, assumed to be solved by retrieval [Blankenberger, 2001; Campbell & Gunter, 2002; Graham & Campbell, 1992; Miller, Perlmutter, & Keating, 1984; Torbeyns, Verschaffel, & Ghesquière, 2002]).

However, Kirk and Ashcraft (2001) questioned the validity of verbal reports for identifying the arithmetic strategies used by individuals. The authors were concerned about the use of this methodology for three interrelated reasons: veridicality, reactivity, and demand. *Veridicality* is related to the accuracy of verbal reports. Kirk and Ashcraft noted that automatic mental processes, such as the retrieval of number facts in memory, are not accessible by verbal reports (Crutcher, 1994; Ericsson & Simon, 1993; Payne, 1994; Wilson, 1994). *Reactivity* has to do with the possibility that the requirement of verbal report may alter the mental processing that normally occurs. For Kirk and Ashcraft, LeFevre et al.'s (1996) participants may have solved the problems atypically because they were expected to report whether they had solved them by remembering or using a procedural strategy. Finally, the *demand-induced bias* concerns the possibility that LeFevre et al.'s results about nonretrieval frequency may have been influenced by instructions that revealed the experimental hypothesis and affected the participants' strategy reports. Smith-Chant and LeFevre (2003) showed, indeed, that participants with relatively low levels of arithmetic fluency respond more slowly and accurately when asked to provide descriptions of their solution procedures. Nevertheless, participants with higher levels of arithmetic fluency show minimal reactivity when asked to provide self-reports. But on our point of view, it is not clear whether or not verbal report requirements had an influence on the strategies used by higher skilled participants. As was mentioned earlier, averaging solution latencies is sometimes misleading, and moreover, a short latency can reflect either a quick procedure or a retrieval of the result from memory.

Therefore, we think, in accordance with Kirk and Ashcraft's (2001) conclusion, that "the use of verbal protocols for simple addition problems is potentially problematic, and great care must be taken in collecting reports for them to qualify as valid reflections of adult performance" and, therefore, that "we must find more appropriate methods of determining the frequency of nonretrieval strategy use among adults" (p. 174).

As a result of (1) the doubts expressed concerning the validity of verbal reports in general and, in particular, the recent questions about LeFevre et al.'s (1996) conclusions and (2) the criticisms directed at interpretations drawn from solution latencies, we propose here a new experimental paradigm that may shed light on the use of the retrieval strategy for arithmetic problems in adults.

This paradigm takes advantage of the fact that algorithmic computation degrades the memory traces of the operands involved in the calculation (Thevenot, Barrouillet, & Fayol, 2001). We have shown, indeed, that the time required for the algorithm to reach the answer and its cognitive cost lead to a reduction in the level of activation of the operands. This decrease in activation is the result both of a memory decay phenomenon, in which memory traces are damaged (Towse & Hitch, 1995; Towse, Hitch, & Hutton, 1998), and of the necessary concurrent activation of transitory results, which induces a resource trade-off or, in other words, a sharing of the attention between the operands, their components, and the intermediary results to be reached to solve the problem (Anderson, 1993). This result was obtained by contrasting the relative difficulty that adults encounter in recognizing the operands after either an addition or a subtraction problem or after their simple comparison with a third number. Let us suppose that an adult has to solve $37 + 28$. It is quite unlikely that he or she will retrieve the result from memory. Different procedures are available, but all of them require the operands to be decomposed (Siegler, 1987). Some adults could decompose 37 and 28 into $30 + 7$ and $20 + 8$, respectively, add 20 to 30, temporarily store the transitory result (50), then compute or retrieve $7 + 8$, and finally add 15 to 50. Other decompositions are possible, but all of them lead to a shift of the attention from the operands (37 and 28) to their components (30 and 7, and 20 and 8, respectively). In contrast, a comparison problem (e.g., decide whether 31 lies between 37 and 28) makes it necessary to keep the numbers in memory without any transformation. Our results clearly showed that the retrieval of two-digit operands from memory is slower and more difficult after their addition or subtraction than after their comparison, which is the behavioral evidence for their lower activation in working memory (Anderson, 1993).

As a consequence, contrasting the relative difficulty that adults encounter in recognizing operands after either their addition or their simple comparison with a third number can allow us to determine whether an addition problem has been solved by an algorithmic procedure or by retrieval of the result from memory. Indeed, if it should prove to have been more difficult to recognize operands after their addition than after their comparison, we can conclude that the operation was solved by an algorithmic procedure. On the contrary, if the difficulty should prove to have been the same in the two conditions, we can conclude that the addition was solved by retrieval—a fast activity that does not imply the decomposition of the operands.

In a first experiment, we studied problems involving three different kinds of numbers: (1) one-digit numbers whose sums were smaller than 10, later referred to as *small numbers*, (2) one-digit numbers whose sums were greater than 10, later referred to as *medium numbers*, and (3) two-digit numbers, later referred to as *large numbers*. If we were to follow LeFevre et al. (1996), the difficulty of recognition of small operands should be the same after their addition as after their comparison with a third number. On the contrary, the difficulty of recognition should be higher after an addition than after a comparison for medium numbers, which were not reported to be solved, in the majority, by a retrieval procedure. As to large operands, they were chosen to maximize the probability of a nonretrieval procedure for their addition. So, we obviously expected to replicate our previous findings (Thevenot et al., 2001): These operands should lead to poorer

recognitions in the addition condition than in the comparison condition. Moreover, we expected this effect to be more pronounced for large numbers than for medium numbers.

In a second experiment, the very same materials as those in the first experiment were used. Moreover, our participants were assessed for their arithmetical skills. Whereas for small and large numbers we expected to observe the same results for low- and high-skilled participants, different patterns of results could be obtained on the problems constructed with medium numbers: Low-skilled participants could, indeed, resort more often to nonretrieval strategies than would high-skilled participants to solve problems of an intermediary difficulty.

EXPERIMENT 1

Method

Participants. Fifty-seven undergraduate students at the Université Blaise Pascal took part in this experiment as volunteers.

Materials. All the experimental trials consisted of four numbers presented sequentially: the first operand, the second operand, an answer, and a target.

The operands consisted of 24 different pairs of numbers (see the Appendix). Eight of these pairs were composed of *small* numbers between 1 and 9, the sums of which never exceeded 10 and the differences of which were larger than one, to allow the comparison with a third number. These 8 pairs were chosen randomly from the set of 16 pairs that fit the previous constraints. Eight pairs were chosen randomly from the set of 12 pairs composed of *medium* numbers between 4 and 9, the sums of which were always superior or equal to 12 and the differences of which were larger than one. Finally, the last 8 pairs were composed of *large* two-digit numbers between 13 and 49, the sums of which fell between 41 and 65 and the differences of which were larger than 10. In order to optimize the probability of an algorithmic procedure, their addition always required a carry and never ended with a 0. These 8 pairs were chosen randomly from the set of 96 pairs that fit the previous constraints.

Each pair of numbers was presented both in the *addition* and in the *comparison* conditions. In the addition condition, the participants had to decide whether a third number corresponded to the sum of the first two numbers, whereas in the comparison condition, they had to decide whether this third number fell between the two numbers previously presented. Each couple of numbers in each condition (addition vs. comparison) was presented twice—once associated with a third number eliciting a *yes* response (the sum of the two operands for addition problems, a number falling between them for comparisons) and once associated with a third number eliciting a *no* response. For addition problems, this erroneous answer was constructed by adding 1 to or subtracting 1 from the correct answer. Indeed, we know that when a proposed answer is too distant from the correct answer, participants are able to solve that kind of problem by approximation, rather than by calculation (Ashcraft & Battaglia, 1978; Ashcraft & Stazyk, 1981; De Rammelaere, Stuyven, & Vandierendonck, 2001; El Yagoubi, Lemaire, & Besson, 2003; Pesenti, Thioux, Seron, & De Volder, 2000). For comparisons, this number was, in half of the cases, superior to the first operand, which was always the largest one, and, in the other half, inferior to the second operand, the smallest.

For the recognition task, each of the experimental trials was associated with a fourth number (i.e., a target) that could correspond to the first operand, the second operand, the first operand plus or minus 1, or the second operand plus or minus 1. Of course, to make the recognition time comparison possible, the target was always the same in the addition and the comparison conditions for a specific number pair.

Each participant was therefore presented with 96 experimental trials: 8 pairs of numbers × 3 sizes of numbers (small, medium, or large) × 2 tasks (addition or comparison) × 2 responses (yes or no).

In order to prevent the participants from adopting a systematic active strategy of memorization of the operands with a view to their subsequent recognition, 192 fillers were added. Those fillers corresponded to the 96 experimental trials without the fourth number. No recognition task was then required from the participants when presented with a filler. Each of the fillers was presented twice. The number of trials that did not require a recognition task was, then, significantly superior to the number of experimental trials. Therefore, it would have been a rather costly and pointless strategy to memorize the operands intentionally and systematically in order to succeed in the recognition task. This precaution was necessary since the extra load in working memory that possibly would have been due to the systematic memorization of the operands could have had an influence on the strategy used by the individuals to solve the problem. Each participant was then presented with 288 trials in a random order.

In order to familiarize the participants with the task, five warm-up trials were presented before the experimental phase.

Procedure. The experiment was controlled by the PsyScope software. The stimuli were presented on screen. Each trial began with the 1-sec presentation of a word that indicated the type of problem to be solved (addition or comparison). This word was replaced by the first number of the pair (first operand). By pressing a key on a button box, the participants removed this number from the screen and displayed the second number (second operand) and then displayed the third one by pressing the same key again. When the third number was displayed on the screen, the participants were asked to give their answer (yes or no) by pressing one of two labeled keys on the button box. The *yes* response was required when the third number either corresponded to the sum of the two first numbers in the addition condition or lay between them in the comparison condition. For the experimental trials only, this response displayed a fourth number on the screen. The participants had to judge whether or not they had seen this number among the first two numbers presented by pressing the same key as that used for the response to the problem. The type of response (yes or no) and the RT were recorded by the computer. This last response displayed a next-trial signal.

For example, an experimental trial with medium numbers might have taken the following form:

Addition / 9 / 4 / 13 / 5,

where the third number (13) required a *yes* response because 9 + 4 = 13, whereas the fourth number (5) required a *no* response because 5 had not appeared previously in this trial. The corresponding filler consisted of the series "Addition / 9 / 4 / 13" only.

Results

The rates of correct responses to the problems were high (.89, .94, and .96 for additions with large, medium, and small numbers, respectively, and .93, .87, and .94 for comparisons), which provides evidence that the participants paid sufficient attention to the problems that preceded the recognition task.

Analysis of the rate of correct responses in the recognition task. Among the 96 experimental trials per participant, only the 48 trials in which a target (first or second operand) was presented in the recognition task were analyzed. Among these 2,736 trials (57 participants × 48 trials), the 176 trials that elicited an incorrect response to the problem were discarded, which was fewer than 6.5% of the trials. A 3 (size of numbers: large, medium, or small) × 2 (type of problem: addition or comparison) × 2 (type

Table 1
Rates of Correct Recognition of the First and the Second Operands As a Function of Their Size and the Type of Problem to Be Solved in Experiment 1

	First Operand				Second Operand			
	Addition		Comparison		Addition		Comparison	
Size	M	SD	M	SD	M	SD	M	SD
Large	.85	.30	.97	.11	.80	.27	.97	.09
Medium	.92	.22	.96	.10	.86	.23	.97	.11
Small	.90	.20	.95	.22	.91	.12	.93	.14

of target: first or second operand) ANOVA with the three factors as repeated measures was performed on the rate of correct answers (see Table 1).

The rate of correct recognition differed as a function of size of numbers [$F(2,112) = 3.16$, $MS_e = 0.02$, $p = .046$]. Recognition was better for small and medium numbers than for large ones (.92, .92, and .90, respectively). Moreover, the rate of correct recognition was lower after an addition (.87) than after a comparison (.96) [$F(1,56) = 13.50$, $MS_e = 0.10$, $p < .001$]. More interesting, there was an interaction between these two variables [$F(2,112) = 7.87$, $MS_e = 0.02$, $p < .001$]. As was predicted, the lower rate of correct recognition for addition than for comparison was observed only for large and medium numbers, but not for small ones [$F(1,56) = 20.91$, $MS_e = 0.06$, $p < .001$; $F(1,56) = 9.62$, $MS_e = 0.04$, $p = .003$; and $F(1,56) = 2.12$, $MS_e = 0.04$, $p = .15$, respectively]. Furthermore, in accordance with our hypothesis, this effect was more pronounced for large than for medium numbers [$F(1,56) = 6.47$, $MS_e = 0.02$, $p = .01$].

Finally, the first operand was recognized more often than the second one [.92 and .91 for the first and the second operands, respectively; $F(1,56) = 4.78$, $MS_e = 0.01$, $p = .03$]. Type of target (first vs. second operand) interacted with type of problem [$F(1,56) = 3.45$, $MS_e = 0.01$, $p = .07$], showing that this effect was obtained only for additions, but not for comparisons [$F(1,56) = 6.03$, $MS_e = 0.02$, $p = .02$, and $F < 1$, respectively].

Analysis of the reaction times in the recognition task. Two hundred two trials on which the target was not recognized were removed from the data set, which was fewer than 8% of the trials. As a consequence, 7 participants for whom there were no more data in one or more experimental conditions were discarded. This analysis was therefore conducted on data from 50 participants. A mean RT was calculated for each participant in each of the 12 experimental conditions: 3 (size of numbers) × 2 (type of problem) × 2 (type of target). An ANOVA with the same design as the previous one was performed on these mean RTs (see Table 2).

The very same pattern of results as that in the previous analysis was obtained. The RTs differed as a function of size of numbers [$F(2,98) = 19.53$, $MS_e = 123,274$, $p < .001$; 1,373, 1,257, and 1,153 msec for large, medium, and small numbers, respectively]. Moreover, the recognition times were higher after an addition problem (1,366 msec) than after a comparison (1,155 msec) [$F(1,49) = 40.90$, $MS_e = 162,716$, $p < .001$]. More interesting, there was

an interaction between these two variables [$F(2,98) = 7.87$, $MS_e = 119,831$, $p < .001$]. In accordance with our hypothesis, the higher RTs for addition problems than for comparisons were observed only for large and medium numbers, but not for small ones [$F(1,49) = 32.31$, $MS_e = 219,983$, $p < .001$; $F(1,49) = 18.03$, $MS_e = 112,371$, $p < .001$; and $F(1,49) = 2.05$, $MS_e = 70,025$, $p = .16$, respectively]. Furthermore, the effect of type of problem was more pronounced for large than for medium numbers [$F(1,49) = 4.84$, $MS_e = 159,513$, $p = .03$].

Finally, the time required to recognize the first operand was shorter than the time to recognize the second [1,216 and 1,306 msec for the first and the second operands, respectively; $F(1,49) = 8.94$, $MS_e = 134,470$, $p = .004$]. However, type of target interacted with type of problem [$F(1,49) = 5.75$, $MS_e = 162,265$, $p = .02$], showing that this effect was obtained only for addition problems, but not for comparisons [$F(1,49) = 8.64$, $MS_e = 246,305$, $p = .005$, and $F < 1$, respectively].

Analysis of the solution times. Although our results confirmed our hypotheses, it was important to show that the better performance in recall for comparisons than for addition problems was not due to longer solution times for the latter, leading to longer delays of recognition and, thus, to a weaker performance. If it were the case, our conclusions would rely only on indirect measures of solution times and would, therefore, be subjected to the criticisms formulated by Siegler (1989) and by LeFevre et al. (1996) evoked in our introduction.

It turns out that solution times (total presentation on screen for the first operand + the second operand + the proposed answer) were, indeed, longer for addition problems (3,904 msec) than for comparisons (3,687 msec) [$F(1,49) = 9.24$, $MS_e = 382,995$, $p = .004$]. However, this difference was due to longer presentations of the second operand for addition problems (1,683 msec) than for comparisons (1,153 msec) [$F(1,49) = 115.31$, $MS_e = 182,9E2$, $p < .001$], whereas the presentation times of the proposed answer (third number) were longer for comparisons (1,402 msec) than for addition problems (1,149 msec) [$F(1,49) = 32.99$, $MS_e = 145,940$, $p < .001$]. The last result was observed whatever the size of numbers [$F(1,49) = 2.82$, $MS_e = 126,595$, $p = .09$; $F(1,49) = 14.62$, $MS_e = 86,330$, $p < .001$; and $F(1,49) = 74.34$, $MS_e = 57,865$, $p < .001$, for large, medium, and small numbers, respectively; see Table 3].

Hence, it is quite unlikely that the differences observed in the rates and RTs for recognition were due to a longer

Table 2
Reaction Times (in Milliseconds) of the First and the Second Operands in the Recognition Task As a Function of Their Size and the Type of Problem to Be Solved in Experiment 1

	First Operand				Second Operand			
	Addition		Comparison		Addition		Comparison	
Size	M	SD	M	SD	M	SD	M	SD
Large	1,476	465	1,164	376	1,647	731	1,205	347
Medium	1,239	389	1,186	379	1,477	570	1,127	301
Small	1,132	349	1,101	278	1,228	426	1,152	307

Table 3
Times of Autopresentation of the First Operand, the Second Operand, and the Proposed Answer As a Function of the Size of the Operands and the Type of Problems to be Solved in Experiment 1

	First Operand				Second Operand				Proposed Answer			
	Addition		Comparison		Addition		Comparison		Addition		Comparison	
Size	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Large	1,185	311	1,187	349	2,500	861	1,234	364	1,389	443	1,508	391
Medium	1,026	263	1,105	316	1,537	490	1,162	305	1,128	322	1,352	394
Small	1,003	325	1,102	285	1,013	300	1,062	287	930	243	1,345	406

retention period in the case of the addition problems. Indeed, it should be remembered that the differences in the rates and times for recognition mainly affected the second operand. It turns out that when this second operand was presented again to the participants for the recognition task, the time elapsed since its first presentation was longer in the case of the comparisons than in the case of the addition problems. Thus, the better recognition of the second operand after a comparison could not have resulted from a shorter period of maintenance. As a consequence, the observed differences between the comparisons and the addition problems cannot be explained only by the shorter solution times for the comparisons.

Discussion

The results of this experiment reveal that the difficulty of recognizing large and medium operands (although to a lesser extent for the latter) is higher after their addition than after their comparison with a third number, which suggests that some of the problems in these two conditions were solved by algorithmic procedures. On the contrary, no such differences were observed between the addition and the comparison conditions when small numbers were involved, which suggests that a retrieval of the answer from long-term memory was the strategy adopted by our participants.

It is crucial to note that the longer recognition times observed for addition problems involving large and medium operands cannot be attributable only to solution times that were longer than those in the comparison condition (see the General Discussion section for an interpretation). Indeed, the time that elapsed between the presentation of the second operand and its recognition was longer in the case of comparisons than in the case of addition problems. The only interpretation of this fact is that the participants calculated the sum of the operands while presented with the second operand and did not wait for the proposed answer to engage in the calculation. As a consequence, when performing the calculation, the individuals had only one number to keep in working memory for a short period of time, which is quite close to an ecological situation in which the numbers are presented simultaneously.

Although both addition problems involving large and those involving medium numbers led to an impairment of the recognition performance, this effect was more pronounced for large than for medium numbers. One possible interpretation of this difference is that some of our participants used a retrieval strategy to solve addition problems involving medium numbers. Indeed, we selected the large numbers in order to trigger nonretrieval procedures for their addition. It is, therefore, quite unlikely that some of the participants solved large problems by retrieval. This could account for the different patterns of results obtained with medium and large numbers: If some of our participants used a retrieval strategy to solve medium problems, it is more probable that they were good at arithmetic, rather than being less skilled. In a second experiment, this hypothesis was examined by taking the arithmetic skills of the participants into account.

EXPERIMENT 2

Method

Participants. Fifty undergraduate students from the Université Blaise Pascal took part in this experiment as volunteers.

The participants were ranked as a function of their arithmetic score on a subset of the French kit (French, Ekstrom, & Price, 1963). In order to maximize the chances to obtain two heterogeneous populations, the 16 participants whose scores were nearest the median were eliminated from the analyses. The analyses were therefore conducted on the data from 34 participants. The median score for the whole group was 50. The *low-skilled* group consisted of 17 participants, with a mean arithmetic score of 41.18 ($SD = 8.85$; range, 19–51). The *high-skilled* group consisted of 17 participants, with a mean arithmetic score of 85.76 ($SD = 15.55$; range, 68–131).

Materials and Procedure. The materials and procedure were the same as those in Experiment 1, except that prior to the operand recognition task, the participants completed both the addition and the subtraction–multiplication subtests of the French kit (French et al., 1963). Each subtest of this arithmetic fluency test consisted of two pages of problems, for a total of four pages. All the participants were given 2 min per page and were instructed to solve the problems as quickly and accurately as possible. The numbers of correct answers on each of the addition and the subtraction–multiplication tests were summed to yield a total arithmetic score.

Results

As in the previous experiment, the rates of correct responses to the problems were high (.91, .95, and .97 for addition problems with large, medium, and small numbers, respectively, and .94, .87, and .96 for comparisons). Overall, the rates of correct responses were quite similar among high-skilled (.94) and lower skilled (.92) participants. This result is not surprising, since the participants were ranked as a function of their score on a French kit subset: The score calculated for each participant corresponded to the numbers of operations that he or she had solved correctly within a fixed period of time. It turns out that the rates of errors in this test were roughly the same for all the participants and that their scores depended mainly on the speed with which they solved the operations.

Table 4
Rates of Correct Recognition for Low- and High-Skilled Participants As a Function of the Size of the Operands and the Type of Problem to Be Solved in Experiment 2

| | Low Skilled | | | | High Skilled | | | |
| | Addition | | Comparison | | Addition | | Comparison | |
Size	M	SD	M	SD	M	SD	M	SD
Large	.91	.11	.94	.11	.92	.16	.96	.07
Medium	.93	.09	.94	.10	.93	.11	.93	.09
Small	.96	.07	.98	.05	.96	.09	.94	.09

Analysis of the rate of correct responses in the recognition task. Among the 96 experimental trials per participant, only the 48 trials on which a target (first or second operand) was presented in the recognition task were analyzed. Among these 1,632 trials (34 participants × 48 trials), the 114 trials that elicited incorrect responses to the problem were discarded, which was fewer than 7% of the trials. A 2 (skill of participants: low or high) × 3 (size of numbers: large, medium, or small) × 2 (type of problem: addition or comparison) ANOVA with the first factor as a between-subjects measure and the last two factors as repeated measures was performed on the rate of correct responses (see Table 4). Significant results were obtained in the previous experiment whatever the type of target (i.e., first or second operand). As a consequence, this factor was not considered as an experimental one in this experiment.

The ANOVA on the rate of correct recognition did not reveal any significant effect.

Analysis of the reaction times in the recognition task. Ninety-two trials on which the target was not recognized were removed from the data set, which was fewer than 6% of the trials. An ANOVA with the same design as the previous one was performed on the mean RTs (see Table 5).

The RTs did not differ as a function of skill of participants ($F < 1$), but they differed as a function of size of numbers [$F(2,64) = 21.11$, $MS_e = 49,340$, $p < .001$; 1,325, 1,135, and 1,093 msec for large, medium, and small numbers, respectively]. The recognition times were higher after an addition problem (1,256 msec) than after a comparison (1,112 msec) [$F(1,32) = 11.92$, $MS_e = 88,047$, $p < .001$]. More interesting, there was an interaction between size of numbers and type of problem [$F(2,64) = 13.75$, $MS_e = 52,666$, $p < .001$]. In accordance with our hypothesis and as in the previous experiment, the higher RTs for additions than for comparisons were observed only for large and medium numbers, but not for small ones [$F(1,32) = 17.06$, $MS_e = 136,348$, $p < .001$; $F(1,32) = 4.11$, $MS_e = 35,036$, $p < .05$; and $F < 1$, respectively].

Moreover, the interaction between the three variables was marginally significant [$F(2,64) = 2.95$, $MS_e = 52,668$, $p < .06$]. For medium numbers, the effect of type of problem was more pronounced for the low- than for the high-skilled participants [$F(1,32) = 5.12$, $MS_e = 35,036$, $p < .04$]. In fact, whereas the RTs were longer for additions than for comparisons for low-skilled participants [$F(1,32) = 9.21$, $MS_e = 35,036$, $p < .005$], they were

identical for high-skilled participants ($F < 1$). However, the effect of type of problem was the same for low- and high-skilled participants concerning small and large numbers ($F < 1$ for both).

Discussion

The results of this experiment show clearly that adults who were less skilled at arithmetic did not solve problems involving medium numbers in the same way as higher skilled adults. Indeed, for this type of problem and for high-skilled participants, the times of recognition for the operands involved in an addition problem were the same as the times of recognition for the same operands involved in a comparison, which suggests that the high-skilled participants relied on retrieval. On the contrary, for the lower skilled participants, the times of recognition for the operands were longer after their addition than after their comparison, which suggests that these participants had to resort to nonretrieval strategies on some of the trials. However, the same strategies were used by the participants to solve problems involving small and large numbers (i.e., a retrieval and a nonretrieval strategy, respectively).

GENERAL DISCUSSION

The main goal of this article was to propose and to test the validity of an original paradigm that could allow us to assess individuals' strategies to solve arithmetic problems without relying on verbal reports or on solution latencies. The conclusions of previous works in which these methods have been used have, indeed, been criticized (see the introduction).

Our results support and extend LeFevre et al.'s (1996) conclusion that adults do not systematically rely on retrieval procedure to solve simple addition problems. We have shown that it is not more difficult to recognize small operands (i.e., single-digit numbers whose sums are lower than 10) after their addition than after their comparison, which suggests that adults solve these very simple addition problems by retrieval of the answer from long-term memory. On the contrary, poorer rates and/or higher times of recognition for operands were observed when large numbers (i.e., two-digit numbers) and medium numbers (i.e., single-digit numbers whose sums are higher than 10) were previously involved in addition problems, rather than in comparisons. This result leads to the conclusion that adults rely on nonretrieval procedures to solve some of these problems.

Table 5
Reaction Times (in Milliseconds) in the Recognition Task for Low- and High-Skilled Participants As a Function of the Size of the Operands and the Type of Problem to Be Solved in Experiment 2

| | Low Skilled | | | | High Skilled | | | |
| | Addition | | Comparison | | Addition | | Comparison | |
Size	M	SD	M	SD	M	SD	M	SD
Large	1,478	454	1,193	298	1,543	722	1,088	319
Medium	1,265	331	1,070	223	1,098	335	1,108	257
Small	1,078	217	1,129	287	1,077	340	1,088	360

Moreover, in the second experiment, individuals whose scores were low on an arithmetic test did not solve addition problems involving medium numbers in the same way as did individuals whose scores were higher. Indeed, whereas the latter's performance of operand recognition was the same after addition problems and comparisons, it took longer for the former to recognize operands after this first, rather than the second, task. Thus, lower skilled individuals rely on nonretrieval procedures to solve some addition problems involving medium numbers, whereas higher skilled individuals are able to rely on retrieval to solve the same problems. This result is interesting because it allows us to better understand what a good arithmetician is: Is a good arithmetician someone who is able to set up back-up strategies quickly and accurately or someone who is able to rely more often on the retrieval of the result from long-term memory? Chronometrical data cannot shed light on this question, but the results of our study can provide some answers. Indeed, we have shown here that the difficulty of recognition of the operands was not due exclusively to a longer retention period in the case of nonretrieval procedures (i.e., the second operand was less well recognized after an addition problem requiring a procedure, rather than after a comparison, even if the time elapsed since the presentation of this operand was longer in the case of the comparison than in the case of the addition problem). Therefore, when one of our participants had difficulty recognizing an operand, it was not due only to the fact that it took him a long time to solve the problem (i.e., memory decay). The only possible alternative explanation is that a part of his attentional resources had been allocated to extra elements, such as the operands' components and the intermediary results to be reached to solve the problem. Consequently, if good arithmeticians were only fast calculators, they would have had some difficulty recognizing the second operand in our experiment after having carried out their calculation. Since it was as easy for them to recognize it after a comparison than after an addition problem involving medium numbers, we can conclude that they retrieved the results from memory.

In conclusion, we have developed a new paradigm that allows us to assess whether or not individuals rely on retrieval or nonretrieval strategies to solve arithmetic problems. The main advantage of this paradigm is to avoid well-known biases associated with solution latencies and verbal report collections. Indeed, we have shown here that the difficulty of recognition of operands is not due exclusively to a longer retention period in the case of nonretrieval procedures. Consequently, our measure is not a trivial indirect measure of solution latencies, which are considered to be too variable to be averaged (LeFevre et al., 1996; Siegler, 1987, 1989). Moreover, the nonreliance on solution latencies allows us to avoid the difficulty of interpreting longer solution times as reflecting either nonretrieval procedures or slower retrieval processes. Furthermore, in addition to the fact that no verbal reports are required from participants, which does away with the reactivity and veridicality biases described by Kirk and Ashcraft (2001), it is worth noting that our participants were not aware of the object of the study. Therefore, it is not possible to cite the demand-induced bias, described by the same authors, to interpret our results. This demand-induced bias has been recently pointed out as well by Campbell and Austin (2002), who noted that there has been considerable variability in estimates of adults' strategies for simple arithmetic (see Campbell & Xue, 2001, for a recent review). According to the authors, these inconsistent results probably reflect differences across experiments in instructions to participants.

In addition, the fact that this paradigm does not rely on any kind of verbal reports makes it particularly suitable for fMRI studies whose aim is to investigate the strategies used by individuals to solve arithmetic problems. Indeed, the difficulty of speaking into the MR scanner comes from the fact that the repositioning of the head, jaw, tongue, and facial muscles during speech leads to distortions and misregistration in the time series MR images. Even if some techniques have been implemented to limit the consequences of these artifacts (see Birn, Cox, & Bandettini, 2004, for a review), a method that purely eliminates them is obviously preferable. Moreover, all brain-imaging studies could benefit from our paradigm. It is, indeed, assumed that the network-like organization in long-term memory for addition problems takes the form of tables that are stored as verbal associations (e.g., "seven plus two is nine"; see Cohen & Dehaene, 2000; Dehaene & Cohen, 1997; Thevenot & Barrouillet, 2006). The brain activation in the cortical regions involved in verbal processes caused by verbal reports can, therefore, interfere with the activation due to the retrieval of verbal knowledge. A paradigm that does not rely on verbal reports is, therefore, advantageous.

As a consequence, the operand recognition paradigm seems highly adapted for studying the strategies used by individuals to solve arithmetic problems. Our future research will therefore extend its use to the other arithmetic operations. Furthermore, the paradigm will be especially suitable for studying the strategies used by children and, a fortiori, mentally disabled children, for whom the use of verbal reports is even more problematic than it is for adults.

REFERENCES

ANDERSON, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

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

ASHCRAFT, M. H. (1995). Cognitive psychology and simple arithmetic: A review and summary of new directions. *Mathematical Cognition*, **1**, 3-34.

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

ASHCRAFT, M. H., & STAZYK, E. H. (1981). Mental addition: A test of three verification models. *Memory & Cognition*, **9**, 185-196.

BARROUILLET, P., & FAYOL, M. (1998). From algorithmic computing to direct retrieval. Evidence for number and alphabetic arithmetic in children and adults. *Memory & Cognition*, **26**, 355-368.

BIRN, M., COX, R. W., & BANDETTINI, P. A. (2004). Experimental designs and processing strategies for fMRI studies involving overt verbal responses. *NeuroImage*, **23**, 1046-1056.

BLANKENBERGER, S. (2001). The arithmetic tie effect is mainly encoding-based. *Cognition*, **82**, B15-B24.

CAMPBELL, J. I. D. (1995). Mechanisms of simple addition and multiplication: A modified network-interference theory and simulation. *Mathematical Cognition*, **1**, 121-164.

CAMPBELL, J. I. D., & AUSTIN, S. (2002). Effects of response time deadlines on adults' strategy choices for simple addition. *Memory & Cognition*, **30**, 988-994.

CAMPBELL, J. I. D., & GUNTER, R. (2002). Calculation, culture and the repeated operand effect. *Cognition*, **86**, 71-96.

CAMPBELL, J. I. D., & OLIPHANT, M. (1992). Representation and retrieval of arithmetic facts: A network-interference model and simulation. In J. I. D. Campbell (Ed.), *The nature and origin of mathematical skills* (pp. 331-364). Amsterdam: Elsevier.

CAMPBELL, J. I. D., & XUE, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General*, **130**, 299-315.

CARR, M., & JESSUP, D. (1995). Cognitive and metacognitive predictors of mathematics strategy use. *Learning & Individual Differences*, **7**, 235-247.

COHEN, L., & DEHAENE, S. (2000). Calculating without reading: Unsuspected residual abilities in pure alexia. *Cognitive Neuropsychology*, **17**, 563-583.

CRUTCHER, R. J. (1994). Telling what we know: The use of verbal report methodologies in psychological research. *Psychological Science*, **5**, 241-244.

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. (2001). Verifying simple arithmetic sums and products: Are the phonological loop and the central executive involved? *Memory & Cognition*, **29**, 267-273.

EL YAGOUBI, R., LEMAIRE, P., & BESSON, M. (2003). Different brain mechanisms mediate two strategies in arithmetic: Evidence from visual event-related brain potentials. *Neuropsychologia*, **41**, 855-862.

ERICSSON, K. A., & SIMON, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: MIT Press, Bradford Books.

FRENCH, J. W., EKSTROM, R. B., & PRICE, L. A. (1963). *Kit of reference tests for cognitive factors*. Princeton, NJ: Educational Testing Services.

GEARY, D. C. (1994). *Children's mathematical development*. Washington, DC: American Psychological Association.

GEARY, D. C., FRENSCH, P. A., & WILEY, J. G. (1993). Simple and complex mental subtraction: Strategy choice and speed-of-processing differences in young and elderly adults. *Psychology & Aging*, **8**, 242-256.

GEARY, D. C., & WILEY, J. G. (1991). Cognitive addition: Strategy choice and speed-of-processing differences in young and elderly adults. *Psychology & Aging*, **6**, 474-483.

GRAHAM, D., & CAMPBELL, J. I. D. (1992). Network interference and number fact retrieval: Evidence from children's alphaplication. *Canadian Journal of Psychology*, **46**, 65-91.

GROEN, G. J., & PARKMAN, J. M. (1972). A chronometric analysis of simple addition. *Psychological Review*, **79**, 329-343.

KIRK, E. P., & ASHCRAFT, M. K. (2001). Telling stories: The perils and promise of using verbal reports to study math strategies. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 157-175.

LEFEVRE, J. A., SADESKY, G. S., & BISANZ, J. (1996). Selection of procedures in mental addition: Reassessing the problem size effect in adults. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 216-230.

LEMAIRE, P., BARRETT, S. E., FAYOL, M., & ABDI, H. (1994). Automatic activation of addition and multiplication facts in elementary school children. *Journal of Experimental Child Psychology*, **57**, 224-258.

McCLOSKEY, M., HARLEY, W., & SOKOL, S. M. (1991). Models of arithmetic facts retrieval: An evaluation in light of findings from normal and brain-damaged subjects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 377-397.

MILLER, K. F. PERLMUTTER, M., & KEATING, D. (1984). Cognitive arithmetic: Comparison of operations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 46-60.

PAYNE, J. W. (1994). Thinking aloud: Insights into information processing. *Psychological Science*, **5**, 241-248.

PESENTI, M., THIOUX, M., SERON, X., & DE VOLDER, A. (2000). Neuroanatomical substrates of Arabic number processing, numerical comparison, and simple addition: A PET study. *Journal of Cognitive Neuroscience*, **12**, 461-479.

SIEGLER, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, **116**, 250-264.

SIEGLER, R. S. (1989). Hazards of mental chronometry: An example from children's subtraction. *Journal of Educational Psychology*, **81**, 497-506.

SIEGLER, R. S. (1996). *Emerging minds: The process of change in children's thinking*. Oxford: Oxford University Press.

SIEGLER, R. S., & CROWLEY, K. (1994). Constraints on learning in non-privileged domains. *Cognitive Psychology*, **27**, 194-226.

SMITH-CHANT, B. L., & LEFEVRE, J.-A. (2003). Doing as they are told and telling it like it is: Self-reports in mental arithmetic. *Memory & Cognition*, **31**, 516-528.

SVENSON, O. (1985). Memory retrieval of answers of simple additions as reflected in response latencies. *Acta Psychologica*, **59**, 285-304.

THEVENOT, C., & BARROUILLET, P. (2006). Encoding numbers: Behavioral evidence for processing-specific representations. *Memory & Cognition*, **34**, 938-948.

THEVENOT, C., BARROUILLET, P., & FAYOL, M. (2001). Algorithmic solution of arithmetic problems and operands–answer associations in LTM. *Quarterly Journal of Experimental Psychology*, **54A**, 599-611.

TORBEYNS, J., VERSCHAFFEL, L., & GHESQUIÈRE, P. (2002). Strategic competence: Applying Siegler's theoretical and methodological framework to the domain of simple addition. *European Journal of Psychology of Education*, **17**, 275-291.

TOWSE, J. N., & HITCH, G. J. (1995). Is there a relationship between task demand and storage space in tests of working memory capacity? *Quarterly Journal of Experimental Psychology*, **48A**, 108-124.

TOWSE, J. N., HITCH, G. J., & HUTTON, U. (1998). A reevaluation of working memory capacity in children. *Journal of Memory & Language*, **39**, 195-217.

WIDAMAN, K. F., & LITTLE, T. D. (1992). The development of skills in mental arithmetic: An individual differences perspective. In J. I. D. Campbell (Ed.), *The nature and origin of mathematical skills* (pp. 189-253). Amsterdam: Elsevier.

WIDAMAN, K. F., LITTLE, T. D., GEARY, D. C., & CORMIER, P. (1992). Individual differences in the development of skill in mental addition: Internal and external validation of chronometric models. *Learning & Individual Differences*, **4**, 167-213.

WILSON, T. D. (1994). The proper protocol: Validity and completeness of verbal reports. *Psychological Science*, **5**, 249-252.

APPENDIX
List of the 24 Pairs of Numbers Used
in Experiments 1 and 2

Small Numbers		Medium Numbers		Large Numbers	
First Operand	Second Operand	First Operand	Second Operand	First Operand	Second Operand
5	3	7	5	28	13
6	2	8	4	35	16
7	1	8	5	36	17
7	2	8	6	38	16
7	3	9	4	39	26
8	1	9	5	43	18
8	2	9	6	43	19
9	1	9	7	49	16