Fast automated counting procedures in addition problem solving: When are they used and why are they mistaken for retrieval ?

Uittenhove, Kim; Thevenot, Catherine; Barrouillet, Pierre Noël

How to cite

UITTENHOVE, Kim, THEVENOT, Catherine, BARROUILLET, Pierre Noël. Fast automated counting procedures in addition problem solving: When are they used and why are they mistaken for retrieval ? In: Cognition, 2016, vol. 146, p. 289–303. doi: 10.1016/j.cognition.2015.10.008

This publication URL: https://archive-ouverte.unige.ch/unige:87926
Publication DOI: 10.1016/j.cognition.2015.10.008

Cognition 146 (2016) 289–303

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Fast automated counting procedures in addition problem solving: When are they used and why are they mistaken for retrieval? ☆

Kim Uittenhove, Catherine Thevenot, Pierre Barrouillet *

Université de Genève, Switzerland

ARTICLE INFO

Article history:
Received 4 November 2014
Revised 14 September 2015
Accepted 11 October 2015

Keywords:
Problem solving
Retrieval
Arithmetic
Association
Numerical cognition

ABSTRACT

Contrary to a widespread assumption, a recent study suggested that adults do not solve very small additions by directly retrieving their answer from memory, but rely instead on highly automated and fast counting procedures (Barrouillet & Thevenot, 2013). The aim of the present study was to test the hypothesis that these automated compiled procedures are restricted to small quantities that do not exceed the size of the focus of attention (i.e., 4 elements). For this purpose, we analyzed the response times of ninety adult participants when solving the 81 additions with operands from 1 to 9. Even when focusing on small problems (i.e. with sums ⩽10) reported by participants as being solved by direct retrieval, chronometric analyses revealed a strong size effect. Response times increased linearly with the magnitude of the operands testifying for the involvement of a sequential multistep procedure. However, this size effect was restricted to the problems involving operands from 1 to 4, whereas the pattern of response times for other small problems was compatible with a retrieval hypothesis. These findings suggest that very fast responses routinely interpreted as reflecting direct retrieval of the answer from memory actually subsume compiled automated procedures that are faster than retrieval and deliver their answer while the subject remains unaware of their process, mistaking them for direct retrieval from long-term memory.

1. Introduction

The associative nature of memory is the object of a large consensus in cognitive psychology. As Anderson (1974) noted, the idea that objects or thoughts that have been experienced in close contiguity become associated in memory (Thorndike, 1922), and that these associations govern the subsequent recollection of these objects or thoughts can be traced back to Aristotle in his essay "On memory and reminiscence". Nonetheless, modern theories went further than Aristotle's insights and no longer view memory as a muddled depository of imprints left by experienced contiguities, but as hierarchically structured systems that store organized bundles of associations (e.g., Anderson, 1974; Anderson, 1993; Collins & Quillian, 1972). These neo-associationist theories also suppose that associations can bind together elements that are not necessarily perceived, but also produced by mental computation (Anderson, 1993). The recurrent solving of a problem is assumed to lead to the association in memory of this problem with its answer, an associative process seen as highly adaptive because it is assumed that directly retrieving answers from memory would provide us with faster and more accurate responses than any algorithmic reconstructive process (Logan, 1988).

This theoretical framework has found one of its most perfect fields of application in the domain of mental arithmetic and simple addition problem solving. Before any systematic tuition in primary school, children develop a variety of counting strategies for solving simple additions. These strategies that initially rely on manipulatives (objects or fingers) become rapidly internalized as verbal counting. Eventually, solving frequently encountered problems by counting procedures leads to their association in long-term memory with the computed answers, adult performance being characterized by the subsequent retrieval of these problem-answer associations. Consequently, development would take the form of a progressive shift from algorithmic problem solving to direct retrieval. The aim of this article is to put this conventional wisdom of cognitive psychology under scrutiny.

1.1. Retrieval of associations in mental arithmetic

A popular application of the associationist framework outlined above is probably the distribution of associations model proposed by Siegler and Shrager (1984). The model distinguishes between

☆ This research was supported by a grant from the Swiss National Science Foundation N° 100014_140257/1 to Pierre Barrouillet and Catherine Thevenot.
* Corresponding author at: Université de Genève, Faculté de Psychologie et des Sciences de l'Education, 40 bd du Pont d'Arve, 1211 Genève 4, Switzerland.
E-mail address: Pierre.Barrouillet@unige.ch (P. Barrouillet).

http://dx.doi.org/10.1016/j.cognition.2015.10.008
0010-0277/

the representation of knowledge about particular problems and strategies that operate on this knowledge to produce responses that in turn modify representations. These representations are conceived as associations of various strength between problems (e.g., 5 + 3) and potential answers that can be correct but also incorrect (e.g., 6, 7, 8, or 9). The determinant dimension of the strategy choice is the *peakedness* of the distribution of associations for a given problem. Some problems have a peaked distribution with an answer, ordinarily the correct answer, that concentrates almost all the associative strength. Other problems have a relatively flat distribution in which the associative strength is distributed among several answers. Retrieving a given problem-answer association within this model depends on three parameters: its relative strength over all the other associations, a confidence criterion that determines the associative strength that must be exceeded for successful retrieval, and a search length criterion that determines the number of retrieval efforts the subject will make before moving to another strategy. The problem is solved through retrieval if an answer is found with an associative strength that exceeds the confidence criterion before reaching the search length deadline. As a consequence, retrieval is more probable for problems with a peaked than a flat distribution.

More relevant for the present study is the assumption of the authors about how children acquire these distributions. In line with the associationist framework, Siegler and Shrager (1984) assume that each time children answer a problem, the associative strength linking this problem to that answer increases, whatever this answer and the strategy used. Thus the probability of retrieval is influenced by the frequency of exposure to the problem, which determines the opportunities to learn answers, and the sum of the two addends, with a greater probability to err when using counting procedures on large numbers. A computational simulation of the model integrating these factors showed that the choice of strategy converges toward direct retrieval, especially for the smallest problems that are more frequently and accurately solved by preschoolers.

This model has received strong support from several studies (Barrouillet & Fayol, 1998; Campbell & Timm, 2000; Geary & Brown, 1991; Geary & Burlingham-Dubree, 1989; Hamann & Ashcraft, 1986; Imbo & Vandierendonck, 2007; Imbo & Vandierendonck, 2008; Reder, 1988) and has provided a theoretical basis to the recurrent observation that adults retrieve from memory the answer of small additions instead of having to calculate it (Ashcraft, 1982; Ashcraft, 1987; Ashcraft & Battaglia, 1978; Ashcraft & Stazyk, 1981; Barrouillet & Fayol, 1998; Campbell, 1987a; Campbell, 1987b; LeFevre, Sadesky, & Bisanz, 1996; Miller, Perlmutter, & Keating, 1984). Thus, it is almost universally admitted that small additions have so often been encountered that their answer is necessarily retrieved from memory in adults (see Zbrodoff & Logan, 2005, for a review).

1.2. A discordant phenomenon: the problem-size effect

A straightforward prediction of the algorithmic computing/direct retrieval transition model would be the progressive attenuation and, at the end, the disappearance of the effects related with factors that affect performance when problems are solved through algorithmic computing. This is the case of the size of the operands in addition solving. In a seminal study, Groen and Parkman (1972) observed that the best predictor of the RTs in first graders asked to solve small additions (the largest problem was 5 + 4) was the size of the smaller of the two addends. This finding suggested the use of a counting procedure by which children start from the larger addend and then count on by ones for the value of the smaller addend (e.g., for 2 + 4, counting 4, 5, 6, a procedure known as the *Min* strategy). The observed slope of 410 ms per increment lent strong support to this hypothesis. Interestingly, tie problems (e.g., 3 + 3) seemed to remain immune to this problem-size effect. Characterized by smaller RTs than the other problems, they were assumed to be solved by direct retrieval of their answer from long-term memory, an idea that is now universally admitted. Groen and Parkman also investigated addition solving in adults. The hypothesis of a transition from algorithmic computing to direct retrieval would have predicted a generalization of the pattern observed in tie problems to all the small problems that were presented in the children study. However, the authors observed a small but significant slope of 20 ms associated with the size of the *Min*. Groen and Parkman judged these 20 ms an implausibly fast rate for a counting procedure, and suggested that adults solve small additions through retrieval, the remaining small size effect being due to the sporadic use of slower counting strategies in rare trials on which the retrieval strategy failed (approximately 5%).

This problem-size effect (i.e., the increase in latencies with the size of the *Min* or the sum of the two operands) has been observed in virtually all the studies, Zbrodoff and Logan (2005) entitling their review on this phenomenon "What everyone finds". The hypothesis of a size effect due to the use of slower non-retrieval strategies in some trials was buttressed by LeFevre et al. (1996) who observed that adults reported using retrieval in more than 80% of the small additions (sum $\leqslant$ 10), but in only 47% of the large additions (10 < sum $\leqslant$ 17) when ties were excluded. However, they also noted that RTs increased with problem size even in those trials that were reported as retrieved. This latter problem-size effect on retrieved small problems was reduced when compared with the effect on all the trials, but somewhat incompatible with the reported process of retrieval. It has nonetheless received several explanations. LeFevre et al. suggested that retrieval latencies could reflect acquisition history, with problems often solved through algorithmic strategies in the course of development resulting in flatter distributions of associations and longer retrieval latencies (e.g., Siegler & Shrager, 1984, contrasted the peaked distribution of 4 + 1 with the flatter distribution of 4 + 5). In the same way, Hamann and Ashcraft (1986, see also Ashcraft & Guillaume, 2009), suggested a memory strength model assuming that the frequency with which additive problems are practiced by children decreases as the size of the operands increases, leading to weaker associations (recall that the frequency of exposure to problems was one of the factors determining the probability of retrieval in the distribution of associations model). Along with this frequency hypothesis, structural properties of the problems have also been advocated. Ashcraft and Battaglia (1978) and Ashcraft and Stazyk (1981), who rejected Groen and Parkman's (1972) hypothesis of a size effect due to the sporadic recourse to slower non-retrieval strategies, suggested that it resulted from the time-course of a search through a tabular representation of the 100 basic addition facts. Beginning at 0,0 and progressing outward along the rows and columns until the intersection is reached, this search would take longer for larger operands. Because the best predictor of response times was the square of the sum in Ashcraft and Battaglia (1978), they hypothesized some stretching of the table in the region of the larger numbers resulting in a slowing down of the search process with larger operands. By contrast, Widaman, Geary, Cormier, and Little (1989), who found that the product of the two addends was the best predictor, hypothesized an equal spacing of the rows and columns of the table from 0 to 9. Assuming a process of spreading activation through the memory network, the time needed to reach a given intersection (i.e., the correct sum) would be proportional to the area of the network to be traversed, hence the predictive power of the product of the two addends. Zbrodoff (1995) and Zbrodoff and Logan (2005) proposed a network interference model in which problem-answer associations take longer to retrieve for larger problems because they suffer from more interference created by overlap of their operands or answers.

In summary, these accounts converge on the idea that the size effect on small additions results from structural or functional characteristics of a process of retrieval from a memory network that stores associations as described in Siegler and Shrager's model (1984). Within this conception, it has even been claimed that the term "problem-size effect" is a misnomer (Ashcraft, 1992), because it remains uncertain that the effect results from the size of the operands per se. This size could coincidentally be predictive of RTs in virtue of its relationships with more central variables like problem difficulty and experience with the problems, which determine the strength of the stored associations and the amount of interference at retrieval. Accordingly, Zbrodoff and Logan (2005) observed that the RTs do not increase monotonically with operands magnitude, as it would be the case if this increase reflected a genuine problem-size effect.

1.3. Recent evidence for a problem-size effect due to counting strategies

Despite a large scientific consensus, some recent studies have called into question the received view that adults solve small additions by direct retrieval. First, Fayol and Thevenot (2012) have observed that the anticipated presentation of the sign for small additions (+) or subtractions (−) 150 ms before the operands leads to faster responses than when sign and operands are displayed simultaneously on screen, an effect that is not observed with multiplications (see Roussel, Fayol, & Barrouillet, 2002, for a related finding). Interestingly, this effect that occurs even with very small problems (e.g., 2 + 3), does not seem to affect tie problems. The authors accounted for this priming effect by suggesting that the anticipated presentation of the sign activates some procedure for additions and subtractions that does not exist for multiplications that are mainly solved by retrieval as additions involving ties. The existence of this type of compacted or compiled and highly automated procedures was already evoked by Baroody (1994), though this proposal was generally neglected.

The hypothesis of automated procedures was reinforced by a recent study in which we had adult participants solving very small additions with operands varying from 1 to 4 (Barrouillet & Thevenot, 2013). This study revealed a quasi perfect monotonic size effect in non-tie problems, with RTs increasing with the size of both the first and the second operand (mean slopes of 16 ms and 23 ms respectively). The best predictor of RTs for non-tie problems was the sum of the two operands ($r = .89$) with a slope of 20 ms. This linear trend is at odds with both the exponential increase predicted by Ashcraft's hypothesis (Ashcraft & Battaglia, 1978; Ashcraft & Stazyk, 1981) and Widaman et al.'s (1989) geometric model that predicts an increase proportional to the product of the two operands. It is also difficult to account for by a hypothesis of exposure frequency. As we noted in Barrouillet and Thevenot (2013), the potential effect of differences in frequency has probably been overestimated in accounting for retrieval times in the domain of mental arithmetic. In the linguistic domain, dramatic differences in word frequency (from about 3000 to 60 per million) result in rather small differences in RTs (i.e., 15 ms in a lexical decision task, Ferrand et al., 2011), whereas the observed difference in RTs between probably very frequent additions such as 2 + 1 and 2 + 4 was higher than 90 ms. Finally, the size effect we observed does not seem to result from interferences in a memory network as Zbrodoff, 1995) suggested. An index we called *Overlap* reflecting the number of problems sharing the same answer as the problem under study was the worst among the eight predictors of RTs that we entered in our analyses ($r = .10$, n.s.). Thus, we concluded that the problem-size effect in these very small additions was better explained by the use of automated fast procedures as hypothesized by Baroody (1994). We suggested that these procedures could consist of scrolling an ordered representation such as a number line or a verbal number sequence (Barrouillet & Thevenot, 2013). For example, solving 3 + 2 would involve moving forward of three and then two steps in the verbal number sequence, the arrival point of the process giving the answer. The speed of this process, which is so fast that it can be mistaken for a direct retrieval by the subjects themselves, could result from the fact that this scrolling would occur within a single focus of attention, the size of which is according to Cowan (2001) limited to four as the size of the operands we used.

1.4. The present study

The studies reported in the previous section called into question the assumption of an inescapable transition from algorithmic computing to direct retrieval in solving the smallest additions, which have probably been practiced thousands of times by educated adults. However, the widespread hypothesis of a solving of frequent and simple additions by direct retrieval of problem-answer associations cannot be jettisoned before a close investigation. Although Barrouillet and Thevenot's (2013) results seemed to point toward the hypothesis of compiled procedures, their study had limitations that make premature any firm conclusion. First, it was assumed, but not verified, that the additions studied would have been considered as solved through direct retrieval by defenders of retrieval models. Of course, these additions involved very small operands (from 1 to 4), but it remains possible that some participants resorted to explicit and slow counting strategies that produced the observed size effect, as Groen and Parkman (1972) suggested. Because verbal reports of the strategies used were not collected, this possibility cannot be rejected. Second, the study was limited to a small subset of the 100 possible additions between one-digit operands (i.e., the 16 problems with operands from 1 to 4). It remains possible that the recursive presentation of the same four numbers in all the problems led participants to adopt a specific strategy to cope with the peculiar demands of the task, thus limiting the generalizability of the results.

Thirdly, and more importantly, it remains to be established whether the linear trend observed in RTs with the increase in size of both operands extends beyond size 4. On the one hand, the hypothesis of a rapid counting procedure to solve simple additions would predict such an extension. Indeed, if the compiled automated procedure that we hypothesize in adults results from an automatization of the algorithmic strategies used by children, there is no reason that it cannot be applied to addends larger than 4. On the other hand, if we were correct in surmising that the speed of this procedure is due to the fact that it processes portions of an ordered representation or representations of quantities that can be held in a single focus of attention, it is possible that the hypothesized compiled procedure cannot operate on larger operands. Convergent evidence suggests that the size of the focus of attention is limited to four, which is also the maximum size of the quantities that can be simultaneously grasped and processed in a single attentional focusing by the subitizing process (Cowan, 2001). If this is the case, this compiled procedure could not process quantities larger than 4, and the nature of the processes involved in solving these larger additions remains an open question. It is possible that adults rely for these larger problems on some memory search of stored associations or other less automatized counting procedures. The pattern of response times for these larger problems should inform us about the strategies used.

Thus, the present chronometric study extended Barrouillet and Thevenot (2013) by presenting a large sample of 90 educated adults with the entire set of the 81 possible additions with operands from 1 to 9, each addition being presented in 6 trials to achieve precise estimates of response times. In order to put the hypothesis of a retrieval of problem-answer associations under close scrutiny, we aimed at investigating problem-size effects on those problems that are usually considered in the literature as being solved by direct retrieval. For this purpose, verbal reports of the strategy used by each participant were collected. However, because it has been observed that the method of verbal report alters the strategies that participants would have spontaneously used (Kirk & Ashcraft, 2001), verbal reports were collected in a separate session administered after the experimental sessions. Our main analyses focused on those participants who reported retrieving the answer of all the very small additions studied by Barrouillet and Thevenot (2013) and almost all the small additions (i.e., with a sum $\leqslant 10$). To get the best possible estimate of the time needed to solve each problem, registered RTs were corrected for the sensitivity of the vocal key, which was assessed in a number-naming task in which participants were only asked to utter each possible answer at signal. Moreover, as in Barrouillet and Thevenot (2013), working memory capacities of each participant were measured. Analysing the RT patterns of individuals varying in working memory capacity could shed light on the nature of the cognitive processes underpinning problem solving. Because the duration of even elementary steps of cognition is affected by working memory capacities (Barrouillet, Lépine, & Camos, 2008), the use of either multistep processes such as the rapid compiled procedure evoked above or single-step processes such as direct retrieval should result in contrasted patterns of size effects in high and low working memory span individuals. The use of multistep algorithmic processes in solving small additions would be revealed by quasi perfect linear increases of mean RTs with the size of the operands, along with pronounced interactions with working memory capacities, individuals with lower capacities being more affected by problem-size effects.

2. Method

2.1. Participants

Ninety undergraduate French-speaking students from the Université de Genève (18 males, mean age: 21 years 6 months, $SD = 4$ years 7 months) received course credit for their participation.

2.2. Material and procedure

The experiment took place in two one-hour sessions administered one week apart. During the first session, participants performed the addition task that was preceded by a number-naming task to control for voice-key sensitivity. During the second session, participants performed an additional block of the addition task with verbal report of the strategy they used. Working memory capacities were assessed through three complex span tasks spread out over the two sessions (a reading span task in the first session and an operation span and a counting span task in the second session).

Addition task – In this task, participants were asked to solve the 81 possible additions of one-digit numbers from 1 to 9, with six trials for each addition for a total of 486 trials. Each trial began by a ready signal (an asterisk) centered on screen for 500 ms, which was immediately followed by an addition displayed horizontally in black on a white background (font characters Déjà vu Serif 38). A voice-key stopped the timer when participants gave their response. This response, which was written down by the experimenter to record accuracy, removed the problem from screen, the next trial beginning after a delay of 500 ms. Participants went through six blocks in which the 81 additions were presented in six different random orders, the order of presentation of these blocks being counterbalanced between participants. An additional block was presented at the beginning of the second session with the same procedure except that participants were asked after each addition to report the strategy they used, following LeFevre et al. (1996) method. Participants were given explanations and examples for the retrieval ("you remember the answer that comes spontaneously to your mind"), counting (the example of $7 + 4$ solved through the *Min* strategy was used for illustration), decomposition (solving $7 + 4$ by adding 1 to $7 + 3$), and transformation (solving a simpler addition and adjusting it to the problem at hand, solving $9 + 7$ by subtracting 1 to the result of $10 + 7$) strategies. After having solved each addition, they were invited to describe the strategy they used and to classify it with the help of the examples given by the experimenter.

Number-naming task – The addition task was preceded by a number-naming task to control for voice-key sensitivity to the 17 possible oral responses from *deux* ("two" in French) to *dix-huit* ("eighteen" in French). Each trial began by a 300-ms ready signal ("$$$") followed by a number from 2 to 18 in word format presented on screen for 1000 ms. Participants were asked to identify this number word and prepare themselves to utter it at the onset of a go signal (an asterisk) that appeared 2000 ms after the offset of the number word. This delay was inserted to get the purest possible estimate of the sensitivity of the voice-key without any contamination by the time needed to read number words or retrieve their phonological image from long-term memory. Reaction times were measured from the onset of the go signal to the onset of the oral response that triggered the voice-key and stopped the timer. Each number word was presented four times in random order.

Working memory span tasks – We used a reading span, an operation span and a counting span task inspired from Kane et al. (2004). In the reading span task, participants were presented with series of 2–5 digits for further recall with three series of each length. Each digit was presented for 1000 ms and followed by a sentence for semantic judgment (e.g., "The new mechanic advised him to check more often the level of altruism"). Participants were asked to read each sentence aloud, to give an oral response about its soundness ("yes" or "no"), and to press the space bar for presenting the next digit. At the end of the series, the word "rappel" (recall) appeared on screen for verbal recall of the digits in correct order. The reading span was the total number of digits in the series perfectly recalled in correct order. The operation span had exactly the same structure, procedure, and scoring method except that the memoranda were letters instead of digits and the sentences were replaced by equations to be verified (e.g., "$(4 + 2) - 1 = 5$"). Finally, the counting span required participants to recall series of 2 to 6 letters, each letter being followed by four dices successively displayed on screen at a rate of one dice every 800 ms for enumeration. The three working-memory tasks significantly correlated with each other (r's > .55, p's < .001). Thus, we added z scores to calculate a compound score taken as an index of working memory capacity.

3. Results

Mean error rate for the total set of addition problems in the six experimental blocks was 2.9%. This rate was positively correlated with the size of the problems ($r = .67$, $p < .01$, between sums and error rates), larger problems eliciting more errors. Participants erred on 0.7% of trials on small additions (sum $\leqslant 10$) and on 4.8% of trials on large additions (sum > 10). Along with incorrect trials, 4.6% of trials were removed from reaction times (RTs) analyses due to voice-key failures. Amongst the RTs on the correctly responded trials, 1.9% that differed from the individual mean by more than three standard deviations were considered as outliers and discarded from the analyses. After removal of errors, outliers, and voice-key failures, five data points, all concerning large problems, were missing when looking at each of the 81 addition problems for all 90 participants. We imputed missing data points by replacing them with the value that could be expected based on the average speed of the participant and the average speed of the missing problem for all other participants. More precisely, we added to the mean RT observed for this problem in the other participants the difference in overall mean response time between this participant and the rest of the sample. The RT analysis was conducted on the remaining 90.9% of trials (corresponding to a total of 39,760 trials). Average individual RTs were subsequently corrected according to sensitivity of the voice key by subtracting to these RTs the deviation to the mean of the naming time corresponding to the answer.

3.1. Overall analysis

Mean corrected RTs for each of the 81 problems are displayed in Table 1. Our results replicate several findings usually reported in the literature. Tie problems elicited shorter RTs than non-tie problems (831 ms and 1156 ms, respectively), $F(1,89) = 239.83$, $p < .01$, $\eta_p^2 = .73$. Large problems involved slower responses than small problems, something true for both tie (892 ms and 781 ms, respectively), $F(1,89) = 105.82$, $p < .01$, $\eta_p^2 = .54$, and non-tie problems (1450 ms and 921 ms, respectively), $F(1,89) = 299.82$, $p < .01$, $\eta_p^2 = .77$. These longer response times on larger problems revealed a strong problem-size effect.

The predictive power on the observed mean RTs of the same predictors as those studied in Barrouillet and Thevenot (2013), which are the predictors traditionally examined in mental arithmetic studies, was investigated. Along with traditional structural predictors (i.e., first and second operands, minimum addend, sum and its square, as well as the product of the addends), we also entered in the equation the percentage of retrieval use reported by the participants. On the entire set of problems as well as when tie problems were excluded, the best predictor of RTs was the rate of reported retrievals, with more frequently retrieved answers corresponding to faster responses (Table 2).

The rates of reported retrievals in our participants were particularly high for tie and small problems (97% and 80% respectively), whereas reported retrievals were rarer for large non-tie problems (24%). These rates were close to those reported in LeFevre et al.'s (1996) study (83%, 83% and 46% for ties, small and large problems respectively), although reported retrievals on large problems were rarer in our sample. Beyond the rate of reported retrievals, the best structural predictor was the size of the sum for the entire set of problems, but the magnitude of the minimum addend when considering non-tie problems only. A stepwise regression was performed on the mean RTs for the entire set of problems and for the non-tie problems with the previously mentioned factors as predictors. These analyses revealed that the rate of reported retrievals accounted for the larger part of variance in RTs for the entire sample of problems, $F = 341.89$, $R^2 = .812$, $p < .01$. The second, third, and fourth predictors to enter the model were the minimum addend, $F = 275.21$, $\Delta R^2 = .064$, $p < .01$, the squared sum, $F = 214.54$, $\Delta R^2 = .017$, $p < .01$, and the sum, $F = 178.91$, $\Delta R^2 = .011$, $p < .01$, respectively. When ties were removed, reported retrieval still accounted for the larger part of variance, $F = 290.5$, $R^2 = .806$, $p < .01$, followed by the minimum addend, $F = 254.12$, $\Delta R^2 = .075$, $p < .01$, the squared sum, $F = 193.3$, $\Delta R^2 = .015$, $p < .01$, and the sum, $F = 174.2$, $\Delta R^2 = .017$, $p < .01$. These findings and the strong relation between fast responses and the reported use of a retrieval strategy corroborate previous studies about the size effect (e.g., LeFevre et al., 1996).

However, when carefully analyzed, the size effect we observed reveals unexpected characteristics that seem at odds with most of the accounts that have been put forward (Zbrodoff & Logan, 2005). As Fig. 1 makes clear, the overall problem size effect in our set of data is mainly due to a sharp difference between large and small problems, but the shape of the effect within these two categories of problems when considered in isolation is quite counterintuitive. Considering large problems, RTs strongly increase from sum 10 to 13, but larger problems do not involve longer RTs. In the same way, when focusing on small problems, size effect seems especially pronounced for the smallest problems (sum from 3 to 7), but disappears from sum 7 onwards.

Moreover, the entirety of the size effect observed in our participants cannot be simply explained by the more frequent recourse to a faster strategy of retrieval in smaller problems as LeFevre et al. (1996) suggested. Two findings contradict this straightforward explanation. First, when considered in isolation, small problems are affected by a strong size effect that cannot be attributed to variations in the rate of reported retrieval. For example, this rate for the small problems with a sum of 3 on the one hand and 10 on the other is exactly the same (85%) whereas these problems strongly differ in their mean RTs (756 ms and 960 ms respectively), $F(1,89) = 160.41$, $p < .01$, $\eta_p^2 = .64$. Hence, when considering small problems in isolation, the rate of reported retrievals is no longer the best predictor (the correlation with RTs just reaching significance, $r = -.25$, $p < .05$, Table 2), while the size of the minimum addend, which is the best structural predictor of RTs for the entire set of non-tie problems, is still the best predictor of RTs. Second, as we will see, even when exclusively focusing on those problems reported by the participants as having been solved by retrieval, a large problem size effect remains. Thus, the problem size effect observed in small problems cannot be attributed to variations in the strategies used to solve them, with the smallest problems more frequently solved through the presumed fastest strategy of retrieval.

To further illustrate this latter point, we can compare the size effect when non-tie problems are reported to be solved either by retrieval or reconstructive strategies (i.e., counting and decomposition strategies, Fig. 2). As it could be expected, the use of reconstructive strategies involves a strong size effect, RTs increasing with the size of the sum with a slope of 67 ms per increment. However, the corresponding slope for retrieved problems is far from being null (40 ms), the functions relating mean RTs to sums presenting striking similarities between retrieved and reconstructed answers. However, the conclusions that can be drawn from this analysis are limited, because the mean RTs reported in Fig. 2 do not subsume the same participants. For example, there are far more participants who reported having solved 7 + 9 than 4 + 2 through reconstructive strategies (89% and 19% respectively) and the participants reporting retrieval on a given problem are not the same as those reporting reconstructive strategies.

Our main research interest was on the small problems the answer of which is assumed to be frequently retrieved from long-term memory. In order to avoid the sample problems evoked above, we concentrated our analyses on a sub-sample of participants who virtually retrieved, or more precisely reported to retrieve, the answers of all the small problems. However, before presenting these analyses, we comment the results obtained with large problems.

3.2. Large problems

We have seen above that large problems were characterized by longer solution times than small problems. Whereas large tie prob-

Table 1
Mean RTs (and standard deviations) for the 81 additions in the entire sample of 90 participants.

First operand	Second operand								
	1	2	3	4	5	6	7	8	9
1	738 (83)	772 (99)	801 (125)	815 (121)	811 (131)	846 (170)	865 (162)	817 (143)	859 (129)
2	740 (97)	750 (99)	886 (184)	917 (213)	976 (207)	1025 (227)	1034 (261)	972 (193)	1059 (227)
3	756 (115)	885 (170)	817 (126)	1071 (390)	1078 (322)	1053 (326)	1075 (274)	1322 (353)	1311 (377)
4	819 (130)	920 (227)	1056 (356)	817 (143)	957 (269)	992 (296)	1543 (459)	1269 (419)	1337 (463)
5	777 (105)	942 (218)	1040 (359)	966 (265)	784 (92)	1231 (461)	1741 (709)	1736 (652)	1337 (463)
6	844 (150)	998 (276)	1033 (327)	1001 (327)	1222 (486)	842 (135)	1613 (604)	1726 (705)	1424 (579)
7	846 (146)	1021 (271)	1003 (336)	1427 (433)	1721 (683)	1692 (636)	831 (143)	1678 (624)	1568 (577)
8	795 (114)	936 (198)	1305 (343)	1238 (384)	1695 (652)	1698 (732)	1765 (667)	976 (244)	1497 (499)
9	847 (128)	1030 (228)	1290 (355)	1311 (363)	1334 (561)	1275 (449)	1525 (618)	1477 (537)	920 (179)

Table 2
Correlations between RTs and different predictors for different sets of problems in the entire sample of participants.

Predictors	Problems				
	All[a]	Non ties[b]	Ties[c]	Small[d] (non ties)	Large[e] (non ties)
First	.49	.52	.87	*.18*	*.12*
Second	.53	.58	_	.35	*.24*
Minimum	.67	.87	_	.80	.62
Sum	.72	.83	_	.58	*.40*
Sum^2	.69	.83	.88	.54	*.38*
Product	.71	.86	_	.80	.49
% Retrieval	−.90	−.90	.59	*−.25*	−.61

Note: All values significant at $p < .01$ except those in italics.
[a] Number of problems was 81.
[b] Number of problems was 72.
[c] Number of problems was 9.
[d] Number of problems was 40.
[e] Number of problems was 32.

lems (from 6 + 6 to 9 + 9) were reported to be solved through retrieval (99%), this strategy was rather rare for the other large problems (24%). However, variations between problems in the rate of reported retrieval could shed light on the peculiarities of the size effect affecting large problems, with a steep increase in RTs for problems with a sum from 11 to 13, followed by a plateau. Fig. 3 reveals a striking parallel between reconstructive strategy use and RTs (compare Figs. 1 and 3). Among the large problems, those with sums of 11 and 12 are less often reconstructed (i.e., more often retrieved) than the others, while there are very few variations in the rate of retrieval for the large problems with sums from 13 to 17. Thus, the pattern of the size effect on large problems illustrated by Fig. 1 can be at least partly explained by variations in the strategies used.

Moreover, other factors contribute to the faster responses on problems with sums 11 and 12. The most frequent reconstructive strategy reported by participants was of decomposition (80% of the problems solved through reconstructive strategies), most often around 10. For example, solving 9 + 3 as (9 + 1) + 2 makes the problem easier by taking advantage of the facility of operations of the form 10 + *n*. It can be assumed that this strategy should be easier with a maximum addend closer to 10 and a small minimum addend easy to decompose, in other words when the difference between the two addends is larger. Accordingly, the mean RTs for decomposition of the large non-tie problems were negatively correlated with the difference between the two addends which was their best predictor ($r = -.78$). It can be noted that the larger differences (i.e., 7 and 6) are concentrated on the problems with smallest sums of 11 (9 + 2; 8 + 3) and 12 (9 + 3), contributing to increase the difference between these large problems and the others. About 15% of the large problems were solved by counting procedures for which the best predictor was the minimum addend ($r = .74$). Finally, the RTs for the large problems solved through retrieval did not exhibit any interpretable trend, as Fig. 4 makes clear. We will see that this muddled pattern strongly contrasts with what is observed with the smallest problems.

3.3. Retrieved small problems

Recall that our aim was to explore the nature of the cognitive processes underpinning the production of answers that are

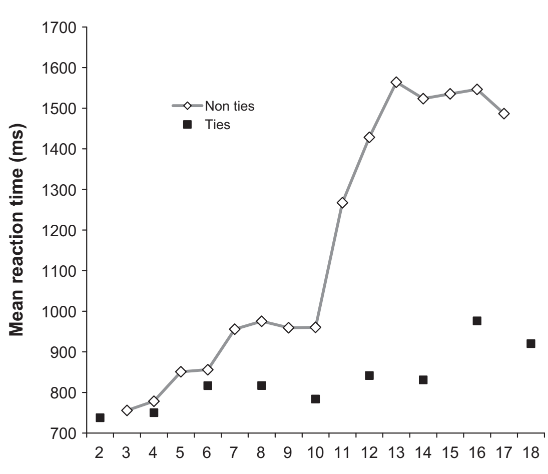

Fig. 1. Mean RTs as a function of the sum for tie and non-tie problems in the entire sample of 90 participants.

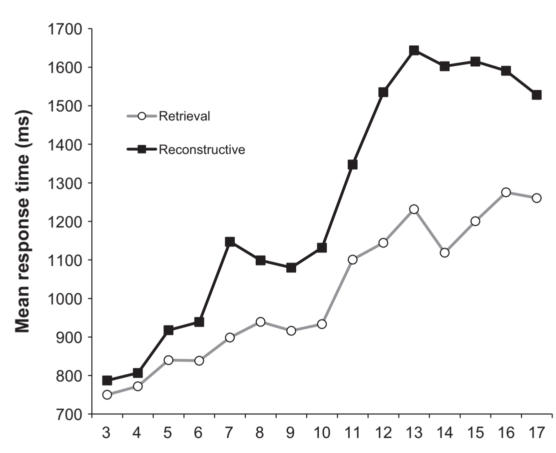

Fig. 2. Mean RTs as a function of the sum for non-tie problems reported as solved by reconstructive strategies or direct retrieval in the entire sample of 90 participants.

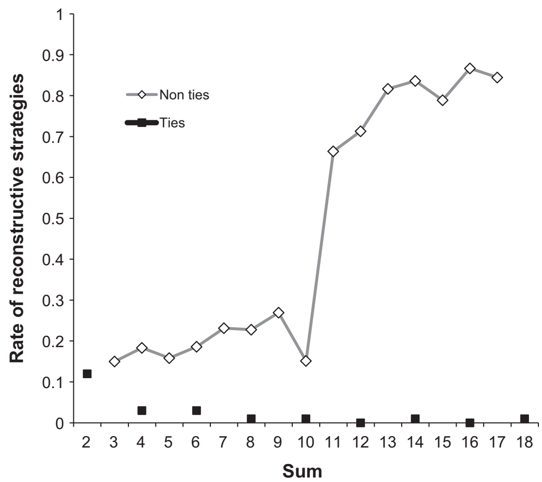

Fig. 3. Rate of reconstructive strategies (i.e., non retrieval) as a function of the sum for tie and non-tie problems.

reported by the participants as having been retrieved from long-term memory. For this purpose, we decided to restrict our analyses of the retrieved small problems to those participants who claimed retrieving at least all the *very small* problems studied by Barrouillet and Thevenot (2013), i.e., the 12 small problems with operands from 1 to 4 when ties are excluded. There were 56 participants in this case. However, we excluded from this sample five participants who had an overall rate of reported retrievals lower than 90% on the entire set of the 45 small problems. This procedure led to the selection of 51 participants who can be considered as frequent retrievers for small problems, with a rate of reported retrievals of 98%. More precisely, apart from the *very small* problems on which these participants were selected with a rate of reported retrieval of 100%, this rate was also 100% on small tie problems (from 1 + 1 to 5 + 5), 98% on the remaining $n + 1$ and $1 + n$ additions (10 problems from $n = 5$ to $n = 9$, hereafter $n + 1$ problems, the problems with $n = 2$ to $n = 4$ being included in the *very small* problems), and 97% on the remaining small problems (18 non-tie problems with sums from 7 to 10 that do not involve 1 with at least one operand larger than 4, hereafter *medium small* problems). The following analyses were run on the mean RTs for the problems reported as retrieved in this sample of 51 participants. Thus, this trimming procedure allowed us to concentrate our analyses on a sample of participants highly coherent on reporting retrieved

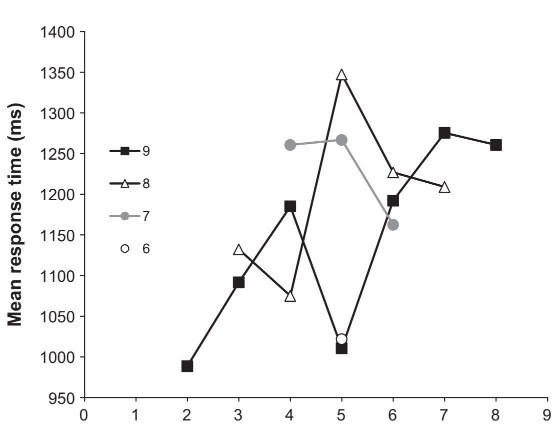

Fig. 4. Mean RTs as a function of the size of the minimum and maximum addends for large problems reported to be solved by direct retrieval.

responses on almost all the small problems. In line with the concordance often advocated between verbal reports and solution times (e.g., LeFevre et al., 1996), these 51 frequent retrievers were faster in solving small problems than the other participants (mean RTs of 860 ms and 966 ms, respectively), $t\,(88) = 3.34$, $p < .01$, and when considering the entire set of 81 additions, their responses were faster when they reported having used retrieval than reconstructive strategies (mean RTs of 907 ms and 1407 ms, respectively), $t\,(50) = 11.73$, $p < .001$. Analyzing size effects on such a sample and on problems reported to be solved by retrieval should provide us with reliable information about the processes underpinning these responses.

First of all, although exclusively focusing on responses reported as retrieved, a strong size effect was nonetheless observable in the 45 small problems with a significant correlation between mean RTs and the size of their sum ($r = .63$, $p < .001$), these RTs ranging from 720 ms for 2 + 1 to 994 ms for 3 + 7. More interestingly, this size effect differently affected the four types of additions we distinguished above (i.e., ties, $n + 1$, *very small* and *medium small* problems). The size effect mainly concentrated on the *very small* problems, whereas there was no size effect at all on the *medium small* problems and only a small effect on ties and $n + 1$ problems (Fig. 5).

In order to assess these size effects, for each of the 51 selected participants, mean RTs per problem were regressed on the size of the sum for each of the four types of additions. Quite counterintuitively, *very small* problems elicited the highest mean slope with an increment of 47 ms per unit ($SD = 37$ ms) that differed significantly from zero, $t\,(50) = 9.14$, $p < .001$. Ties exhibited lower slopes (mean = 8 ms, $SD = 7$ ms) that nonetheless significantly differed from 0, $t\,(50) = 8.87$, $p < .001$, as well as $n + 1$ problems (mean = 7 ms, $SD = 22$ ms), $t\,(50) = 2.26$, $p < .05$. However, there was no size effect on *medium small* problems (mean = −5 ms, $SD = 44$ ms), $t\,(50) = -0.74$, $p > .20$. The slope for the size effect associated with the *very small* problems was significantly steeper than for the three other categories of small problems, $ts\,(50) > 5.85$, $ps < .001$, whereas ties and $n + 1$ problems did not differ from each other, $t < 1$, and had slopes slightly steeper than *medium small* problems, $t\,(50) = 2.07$, $p < .05$, and $t\,(50) = 1.95$, $p = .06$, respectively.

It seems rather difficult to interpret the small but significant size effect observed on tie problems, which seems mainly due to very fast responses on 1 + 1 and 2 + 2. It is worth to note that the slope associated with the small tie problems was not greatly affected when also taking into account large tie problems up to 9 + 9. Whereas the slope was of 8 ms for the small tie problems, it only increased to 12 ms when all the tie problems were taken into account. In the same way, the irregular pattern that gives raise to the small size effect that affects $n + 1$ problems remains unclear. Note that it cannot be due to a differential sensitivity of the voice key to the utterance of the different answers because this potential source of variation was controlled in the present experiment. These effects contrast with the clear size effect observed in *very small* problems. It is worth to note that the larger size effect that affected *very small* problems compared with *medium small* problems was not due to the fact that very small problems included some problems known to elicit fast responses such as those involving 1. When additions involving 1 (i.e., 2 + 1, 1 + 2, 3 + 1, 1 + 3, 4 + 1, and 1 + 4) were removed from the *very small* problems, their slope remained highly significant (34 ms, SD = 61 ms), $t\,(50) = 4.06$, $p < .001$, and still steeper than the slope associated with the *medium small* problems, $t\,(50) = 3.87$, $p < .001$. Interestingly, the *very small* problems involving 1 exhibited themselves a steeper slope than what we call the $n + 1$ problems (i.e., with $n > 4$, mean slopes of 28 ms and 7 ms, respectively), $t\,(50) = 3.18$, $p < .001$. In other words, there is a size effect intrinsically related with the problems

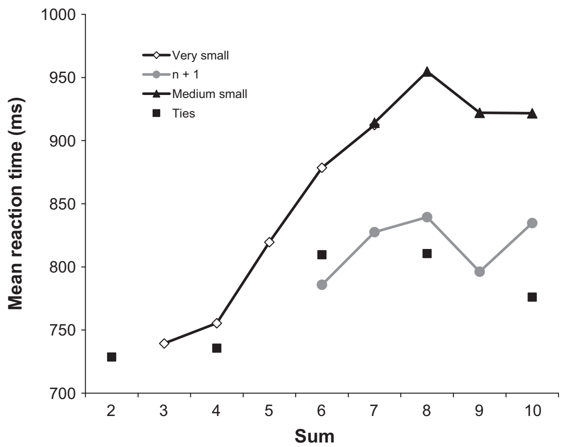

Fig. 5. Mean RTs for the four types of retrieved problems in a sub-sample of 51 participants classified as frequent retrievers (for the categories *very small*, *n* + 1, *medium small*, and ties, see text).

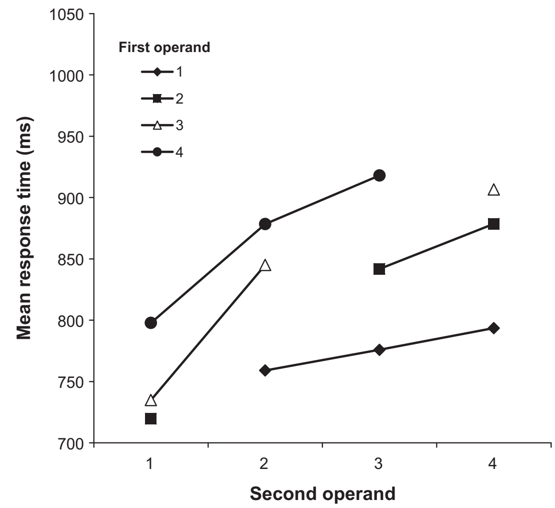

Fig. 6. Mean RTs for the *very small* problems as a function of the magnitude of the first and second operands in the sub-sample of the 51 participants classified as frequent retrievers.

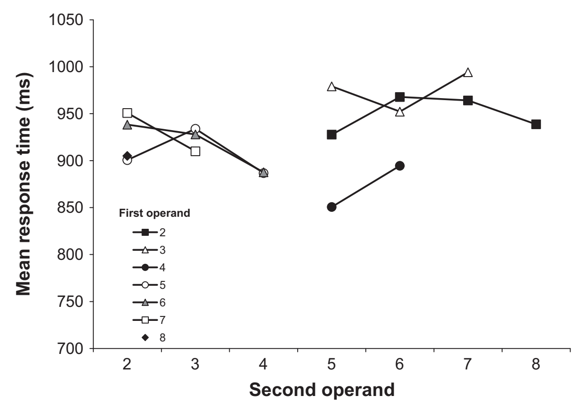

Fig. 7. Mean RTs for the *medium small* problems as a function of the magnitude of the first and second operands in the sub-sample of the 51 participants classified as frequent retrievers.

involving operands that do not exceed 4 (i.e., the *very small* problems) that is significantly stronger than the size effect that can affect any other type of small problems.

Thus, the present study totally confirms the results of Barrouillet and Thevenot (2013) and the size effect they observed on *very small* additions. As in this previous study, the RTs for *very small* problems were highly correlated with the sum of the two operands that was their best predictor ($r = .95$). As shown in Fig. 6, RTs monotonically increased with the size of both the first and the second operand.

The slopes related with the increase in size of the second operand were of 17 ms, 54 ms, 53 ms, and 60 ms when the first operand was 1, 2, 3, and 4 respectively for a mean of 46 ms, whereas the slopes related with the increase in size of the first operand were of 39 ms, 40 ms, 46 ms and 57 ms when the second operand was 1, 2, 3, and 4 respectively, for a mean of 45 ms. This pattern strongly suggests some sequential mechanism taking about 50 ms per increment, whatever the incremented operand. This is in sharp contrast with the *medium small* problems. Fig. 7 displays their mean RTs in the same way as for the *very small* problems (i.e., as a function of the size of both the first and the second operand). No clear trend appears, except that *medium small* problems take longer to solve than the *very small* problems with a mean RT of 921 ms that corresponds to the RTs associated with the slowest *very small* problems (907 ms and 918 ms for 3 + 4 and 4 + 3 respectively).

3.4. Individual differences on retrieved small problems

As we argued above, the analysis of individual differences can shed light on the nature of the cognitive processes by which answers are produced. Assuming that individual differences related with working memory on a given task reflect the concatenation of differences elicited by its elementary constituents (Barrouillet et al., 2008), these differences should be more and more pronounced as the number of steps involved in the process under study increases. Thus, size effects resulting from the increased number of steps a multistep process involves should be especially affected by differences in working memory capacities. Among the 51 participants previously classified as frequent retrievers, and when focusing on problems reported as retrieved, mean RTs on the small problems were negatively correlated with working memory capacities ($r = -.49$, $p < .01$), the higher these capacities, the faster the responses. This correlation was observed in each of the four types of small additions we distinguished above (r values of −.47, −.48, −.44, and −.47 for ties, *very small*, *n* + 1, and *medium small* problems, respectively). More interestingly, the slope associated with the size effect (i.e., the sum) on the small problems was also negatively correlated with working memory capacities ($r = -.36$, $p < .01$), indicating that the lower these capacities, the steeper the slope and the stronger the size effect. However, when considering the four types of small problems that we distinguished above, this correlation was only significant for the *very small* problems ($r = -.43$, $p < .01$), whereas it did not reach significance in any of the other types of problems (r values of −.14, .15, and .11 for ties, *n* + 1, and *medium small* problems, respectively, $ps > .10$). To illustrate this point, we contrasted the third of the subsample of frequent retrievers (17 participants hereafter referred to as high-span participants) who had the highest compound working memory scores with the third who achieved the lowest scores (hereafter referred to as low-span participants). High-span participants were faster than low-span participants in solving small problems (mean response times of 779 ms and 936 ms respectively), $t(32) = 4.42$, $p < .001$, and exhibited a smaller size effect (mean sum-related slopes of

18 ms and 33 ms, respectively), t (32) = 3.35, $p < .01$. The difference in slopes was especially pronounced for *very small* problems (mean slopes of 29 ms, SD = 22 ms, and 70 ms, SD = 36 ms, respectively), t (32) = 3.97, $p < .001$. Whereas both groups exhibited a significant size effect for tie problems, with mean slopes (9 ms and 11 ms for high- and low-span individuals, respectively) that differed significantly from 0, t (16) = 6.17 and 8.28, respectively, $ps < .001$, these size effects did not significantly differ between groups, t (32) = 1.40, $p > .10$. In the same way, there was no significant difference between the two groups in the slopes associated with the sum of the $n + 1$ problems (11 ms and 9 ms for high- and low-span individuals, respectively) or the *medium small* problems (7 ms and −4 ms, respectively). Apart from the $n + 1$ problems in high-span participants, t (16) = 3.21, $p < .01$, none of these size effects reached significance. Interestingly, the difference in size effect between the high- and the low-span groups was significantly larger for *very small* than tie problems, as testified by the significant interaction between groups and types of problems revealed by an ANOVA on the slopes relating RTs to the sum of the problems with the type of problems (*very small* vs. ties) as within-subject factor and working memory capacities (high vs. low) as between-subject factor, F (1,32) = 13.43, $p < .001$, $\eta_p^2 = .30$ (Fig. 8).

When considering the *very small* problems, the pattern of RTs was similar for high- and low-span individuals. In both groups, RTs increased with the size of the first and the second operand. In high-span individuals, the slopes related with the increase in size of the second operand were of 6 ms, 32 ms, 32 ms, and 47 ms when the first operand was 1, 2, 3, and 4 respectively for a mean of 30 ms, whereas the slopes related with the increase in size of the first operand were of 17 ms, 24 ms, 35 ms and 32 ms when the second operand was 1, 2, 3, and 4 respectively, for a mean of 27 ms. The pattern was the same in the low span group, except that the slopes were far steeper (20 ms, 76 ms, 82 ms, and 97 ms for a mean of 69 ms concerning the increase in size of the second operand; 50 ms, 59 ms, 71 ms, and 90 ms for a mean of 67 ms concerning the increase in size of the first operand). These findings suggest that the sequential process that underpins the increase in RT with the size of both operands of *very small* problems depends on working memory, this process being faster in individuals with higher working memory capacities.

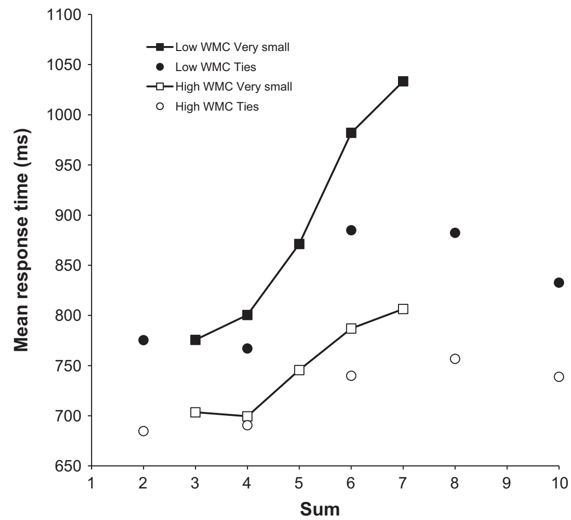

Fig. 8. Mean RTs for the *very small* problems and ties a as a function of the sum for individuals with low and high working memory capacities (WMC).

4. Discussion

The aim of this study was to test the commonly-held assumption that the recurrent solving of problems through counting strategies leads to associate these problems with their answer in memory, and that, with practice, these associations become strengthened to the point that their retrieval is faster than any other strategy, leading to a developmental shift from algorithmic solving to direct retrieval from memory. These key assumptions were tested in the domain of additive problems solving. From a sample of 90 undergraduate students who solved the 81 additions from 1 + 1 to 9 + 9, we selected a sub-group of 51 participants who reported using retrieval in almost all the small problems (i.e., mean rate of 98%) and, for each of these frequent retrievers, we only retained those problems he or she reported as having solved by retrieval. The mean RT for small problems in this sub-group was 857 ms, far smaller than the mean RT for problems they reported having solved by some algorithmic procedure (1048 ms), and also smaller than the mean RT exhibited by less frequent retrievers when solving small problems (i.e., 966 ms in the 39 remaining participants). Thus, the database that we analyzed presents all the characteristics that are usually taken as evidence for direct retrieval. Nonetheless, the observed pattern of RTs is at odds with the hypothesis that all these problems were solved by direct retrieval of their answer from memory. Three main findings contradict this belief.

First of all, although our analysis focused on trials that exhibited all the characteristics usually considered as reflecting a process of retrieval, a strong size effect was observed in small problems, with mean RTs on non-tie problems ranging from 720 ms to 994 ms. It is difficult to imagine that a one step process like direct retrieval would exhibit such variations in duration. More intriguingly, this size effect was restricted to the smallest problems, those we called *very small* problems, with operands from 1 to 4. The slope associated with the sum of these problems was steeper than the slope for any other type of small problems we distinguished (ties, $n + 1$, or *medium small* problems). This finding contradicts the retrieval-based theories that account for the size effect on presumably retrieved problems by evoking acquisition history (LeFevre et al., 1996; Siegler & Shrager, 1984), frequency of exposure (Hamann & Ashcraft, 1986), or interference (Zbrodoff & Logan, 2005). Indeed, *very small* problems are precisely the first to be solved by children, the most frequently encountered, and those that involve the lowest amount of interference. Second, and contrary to Zbrodoff and Logan (2005) who claimed that there is no genuine problem-size effect in mental arithmetic because RTs do not increase monotonically with operand magnitude, we observed an almost perfect monotonic increase in RTs with the size of both the first and the second operands (Fig. 6). Each increment in both operands resulted in an increase in RT of about 45 ms. This finding replicates what Barrouillet and Thevenot (2013) observed in another sample of adults when only presented with the *very small* additions. Third, this size effect and the slope associated with the increase in magnitude of the operands of *very small* problems appeared to be related to working memory capacities, something that was not observed for the other types of small problems such as ties, $n + 1$, or *medium small* problems. This finding strongly suggests that *very small* problems are solved by some sequential multi-step process that differs in nature from the process underlying tie, $n + 1$, or *medium small* problem solving.

Overall, our findings are difficult to reconcile with the hypothesis that when solving the *very small* problems, which are the easiest, the most frequently solved, the first to be practiced in preschool age, those that involve the smallest operands and elicit the fastest responses, educated adults retrieve answers from associations stored in long-term memory.[1] An alternative account is clearly needed. Our hypothesis is that these answers are not retrieved but reconstructed through a rapid sequential procedure the duration of which is determined by the magnitude of the operands. Interestingly, this rapid procedure seems to be limited in its application to very small operands that do not exceed 4. Beyond this limit, a variety of mechanisms might contribute to produce responses. We have seen that solution times to *medium small* problems do not increase with the size of the operands, RTs varying in an uninterpretable way from problem to problem (Fig. 7). This insensitivity to the size of the operands is fully compatible with a process of retrieval of the answers from long-term memory, as it was initially assumed by Groen and Parkman (1972) who suggested that the time for retrieving answers should be independent of the problem. In the same way, since Groen and Parkman (1972), it is usually assumed that ties are solved by retrieval (e.g., Campbell, Chen, & Maslany, 2013; Campbell & Gunter, 2002; Fayol & Thevenot, 2012). Our results corroborate this common assumption. Ties exhibited a small size effect, with a slope significantly lower than for *very small* problems, even when the analysis was extended to large problems from 6 + 6 to 9 + 9. It is also commonly assumed that $n + 1$ problems can be solved using a rule consisting of producing the number after n in the counting sequence (Sokol, McCloskey, Cohen, & Aliminosa, 1991). We have seen that whereas problems involving 1 were affected by a strong size effect with a slope of 28 ms when belonging to the *very small* problems (i.e., with n varying from 2 to 4), this slope dropped to 7 ms when n varied from 5 to 9, suggesting that two different strategies are used depending on the magnitude of n. Nonetheless, the slope of 7 ms is totally compatible with the hypothesis that what we called $n + 1$ problems (i.e., with $n > 4$) were solved using such a one-greater rule.

However, the hypothesis that *very small* problems are solved by a rapid sequential procedure whereas *medium small* and ties problems are solved through retrieval and $n + 1$ problems with a rule raises three questions. The first concerns the nature of this rapid procedure that is faster than a retrieval process or a rule. It is actually so fast that participants mistake it with retrieval, answer popping to their mind while they remain unaware of the process itself. The second concerns its formation. It is usually assumed, as in Siegler and Shrager (1984), that the repeated practice of algorithmic computing has its effects in the storage and reinforcement of problem-answer associations, and not in the increased automaticity of the algorithm, the finishing time (distribution) of which is assumed to "stay the same while the finishing time for the retrieval process decreases" (Logan, 1988, p. 496). The third concerns its limitations. This rapid procedure seems to be limited to small numbers up to four, the cognitive system resorting to other strategies like retrieval or rules only when this limit is exceeded. In the following, we address these three questions in turn.

4.1. Rapid automatized procedures

The hypothesis that, after extensive practice, the answer of small additions is delivered by highly automated and compiled procedures was first introduced by Baroody (1983, 1984, 1994). He suggested that the key change in number fact efficiency does not result from a shift from slow counting procedures to memory retrieval, but rather from slow to automatized and faster procedures. He stated "as the child learns rules, heuristics and principles, these supplant less efficient procedural processes such as informal counting algorithms. Moreover, as these rules, heuristic and principles become more secure and interconnected, their use becomes automatic. As a result, problem solving becomes more efficient" (1983, p. 227). However, Baroody did not precisely describe these automated procedures, and his hypothesis remained in neglect. Nonetheless, fine-grained descriptions of this type of automated and fast cognitive process have been proposed, such as the *decision cycle* described by Newell (1990) in his Soar model. Newell distinguishes four levels of cognition from productions to decision cycles, primitive operators and finally goal attainment. We will concentrate on the first two levels. The lowest level is made, in Soar system, of productions relating conditions to actions. Conditions correspond to elements in working memory, whereas actions enter into the working memory new elements corresponding to encoded knowledge in long-term memory. In some sense, all the long-term memory in Soar can be seen as a single production system acting as a recognition system. From the elements currently held in working memory, which constitute retrieval cues, productions recognize patterns of knowledge in long-term memory and respond by providing the content of these patterns that enter working memory. These productions that access long-term memory and retrieve information from it are assumed to be involuntary and very fast, of the order of tens of milliseconds.

The next level, the decision cycle, corresponds to the smallest deliberate act as well as the smallest unit of serial operations. Basically, the decision cycle accumulates knowledge for act by repeated accessings of knowledge (productions) during an elaboration phase and decides. Its duration is of the order of hundreds of milliseconds and depends on the number of productions involved. Interestingly, the decision cycle is involuntary and automatic. It runs to quiescence and delivers the response, the subject being only aware of the product of the decision cycle, not its process. An example of a process happening within a decision cycle is, according to Newell (1990), that underpinning the response in a basic Sternberg task in which a sequence of target digits is presented to the subject followed by a probe, the task consisting of deciding whether the probe was a member of the set or not (Sternberg, 1966). Typically, response times in this task increase linearly with the size of the set for both positive and negative responses, with a slope of about 40 ms. A finding that has attracted interest for decades is that response times do not depend on the serial position of the probe when present in the set, indicating that the task is not performed using some terminating search that would stop as soon as the probe is encountered. In a nutshell, Soar accounts for this task by assuming that a production is sequentially instantiated that compares each target with the probe. Because the process is automatic, the decision cycle runs to quiescence and delivers a response at the end of the cycle ("yes, the probe is member of the set", or "no, it is not"). Soar accounts for several phenomena related with this task in the following way. The linear trend results from the sequential instantiation of the production. The process is fast (a slope of about 40 ms) because it happens at

[1] It could be argued that recent neuroscientific investigations such as Qin et al. (2014) who observed hippocampal-neocortical reorganization related with the developmental shift from procedural- to retrieval-based addition solving lend strong support to the hypothesis of addition solving through direct retrieval. Despite their interest for brain functioning and reorganization over development, this type of findings could be less constraining than usually assumed for theories of cognitive arithmetic. First of all, Qin et al. (2014) did not aim at demonstrating that there is a procedural-retrieval shift in development. They take this shift for granted based on verbal reports and RT analyses, their investigations and the selection of ROI being oriented by this postulate. However, we have seen that verbal reports and fast RTs are not necessarily reliable indices of genuine memory retrieval. Second, any procedural rule including the compiled procedure we hypothesize relies on memory retrieval (see Fig. 9). Thus, identifying brain activities aligned with memory processing could not be necessarily indicative of memory retrieval of the answer. Third, more than a higher activation of mnemonic systems like the hippocampus, what Qin et al. (2014) have observed is that an increased hippocampal connectivity with prefrontal-parietal circuits predicted longitudinal gain in retrieval fluency. This type of connectivity is exactly what is needed for efficiently running procedures that update working memory content with retrieved knowledge from long-term memory. Finally, we do not deny that the answer of many simple additions is retrieved (e.g., the *medium small* and some large additions), we simply argue that some small additions could be solved by faster automated procedures.

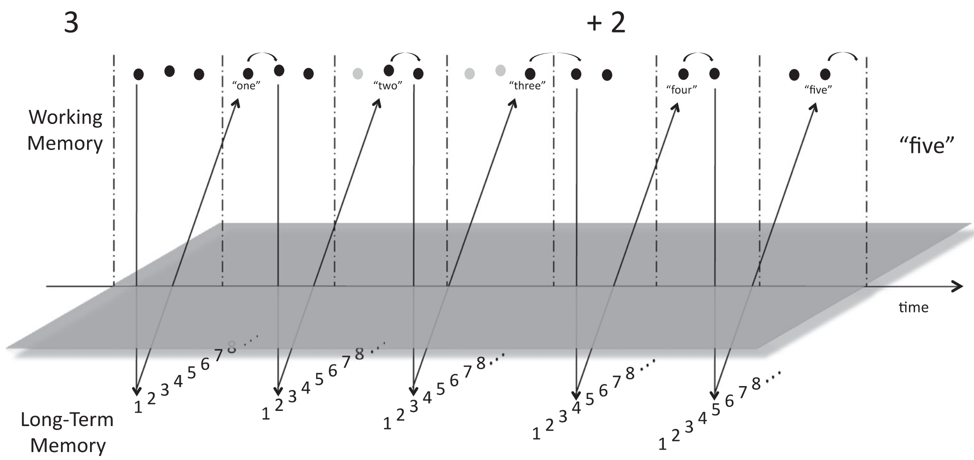

Fig. 9. Hypothetical timeline of events during the implementation of the automatized procedure for solving additions involving small numbers from 1 to 4. The first operand is encoded in working memory in an analogical representation that triggers a *next-token-next-value* production (represented by solid lines) that accesses the number chain in long-term memory, the retrieved numerical value entering working memory and tagging the corresponding token. The production is successively fired by each element of the representation of the first and then the second operand, running into quiescence and delivering the answer. The time elapsed between two successive vertical dashed lines corresponds to a production cycle and is assumed to be of the order of tens of milliseconds. The process is limited to small quantities that can be represented analogically in a single focus of attention (i.e., no more than four elements). It is so fast that the subject is only aware of its product, hence the subjective experience of a direct retrieval from memory.

production rate (i.e., some tens of ms). Response times do not depend on the serial position of the probe when present because the decision cycle is automatic, running to quiescence before delivering a response. Finally, people remain unaware of the process itself because it takes place within a decision cycle.

The automated procedure that we assume underlying *very small* additions solving shares several characteristics with Soar's decision cycle. The linear trend relating RTs to the sum of the operands might result from the sequential instantiation of a production, with the number of instantiations being determined by the magnitude of the operands. The procedure is fast, the slope of the linear function (i.e., about 45 ms) falling within the range of production rate. Like the decision cycle, the procedure is automatic, running to quiescence out of the control of the subject who remains unaware of the process itself, having only access to its outcome. This is why participants so often report having retrieved the answer from memory. Our results do not allow definite hypotheses about the exact nature of the procedure itself, but only speculations illustrated in Fig. 9. Taking into account that the process seems to be limited to small numbers up to four, a point that we will address below, it could be imagined that both operands are successively encoded using some analogical representation that captures the meaning of the numbers to be added. This type of semantic representation could be privileged by the subject when focusing on cardinality, pursuing the goal of manipulating numerical values for sake of transformation in arithmetic operations. In Fig. 9, the analogical representation of the first operand triggers the *next-token-next-value* production that recursively accesses knowledge related with the number chain stored in long-term memory, successively tagging each token within the representation with the next numerical value retrieved from long-term memory. This procedure would successively scan the representation of the first and then the second operand, running to quiescence and delivering the last accessed value as a response. Response time would be a function of the number of production instantiations, which corresponds to the sum of the two operands, resulting in a linear trend with a slope reflecting production rate (about 45 ms in our experiment).

4.2. Acquisition

The fact that *very small* additions are solved by a sequential multi-step process, as the linear size effect that we observed testifies, suggests either that extended practice does not necessarily result in associating problems with their answer in long-term memory or, if such associations do exist, that their retrieval is slower than this multi-step process that consequently determines performance. This raises the question of the acquisition of this kind of automated procedure. Baroody (1983) suggested that the slow informal counting algorithms used by young children are progressively supplanted by rules, heuristics, and principles, the use of which becomes automatic. One of the main characteristics of these counting algorithms is that young children use their fingers or objects for modeling the problems (Carpenter & Moser, 1983), these manipulatives being progressively internalized, resulting in verbal counting. Of course, these algorithms are at the beginning slow, requiring control and awareness of each of their steps. The traditional associationist approach (e.g. Ashcraft & Fierman, 1982; Siegler & Shrager, 1984), which assumes that automation results from the progressive abandonment of these algorithms for the retrieval of problem-answer associations, is actually based on two questionable assumptions. The first is that any algorithmic solving necessarily results in the association in memory of the problem with the answer obtained. However, Thevenot, Barrouillet, and Fayol (2001) showed that this could not always be the case, especially when algorithmic solutions involve important delays between operands encoding and answer production, as it is the case in children's problem solving. Long delays between operands encoding and response lead to the degradation of the memory trace of the operands and could prevent their association with the answer. The second questionable assumption is that algorithmic solving reinforces operand-answer associations more than the algorithm itself. We saw for example that Logan (1988) assumes that the distributions of the finishing times of algorithms stay the same through practice. It might alternatively be assumed that algorithmic solving primarily reinforces the algorithm itself, its implementation becoming faster and faster through a process of compilation as described by Anderson (1983) and Anderson (1993) in his ACT model. The overt enumeration of each manipulative would be progressively compiled into the recursive *next-token-next-value* production described above, that no longer needs verbalizing of each counting step, only the output being accessible to consciousness.

The automation of such a compiled procedure could be a rather slow and delayed developmental process. We hypothesized that the procedure operates on analogical representations of the quantities the operands refer to (Fig. 9). This means that a prerequisite for the automatic triggering of this procedure is the capacity to automatically convert digits into their analogical representations. However, it has been shown that, although the Arabic code is easily learnt and used by young children, the automatic access to analogue representations from digits does not seem to occur before age 9 or 10 (Girelli, Lucangeli, & Butterworth, 2000; Rubinsten, Henik, Berger, & Shahar-Shalev, 2002). This would explain why the developmental shift from algorithmic computing to direct retrieval has been described as occurring around these ages (Ashcraft & Fierman, 1982).

It could be argued that small addition solving is too fast a process to be carried out by a multistep procedure, as compiled as it may be. However, other numerical processes reach the same speed as the additive procedure with, intriguingly, approximately the same rate. This is the case for subitizing, the capacity to apprehend and enumerate small sets of items up to 4 in an effortless and perfectly accurate way (Kaufman, Lord, Reese, & Volkmann, 1949). Although it exhibits some variability from one study to another, the rate of this process varies in the range of the production rate evoked above (from 25 ms to 60 ms in Klahr & Wallace, 1976). Interestingly, Klahr (1973) associated the subitizing slope with the rate of memory scanning in the Sternberg task, and proposed an account for this slope germane to the *next-token-next-value* production we hypothesize, assuming that it comes partly from the need to map items onto memory representations for number names. Of course, the limitations of the subitizing process and its slope have received various explanations (e.g., Cowan, 2001; Gallistel & Gelman, 2000; Trick, 1992), and its existence has even been called into doubt (Balakrishnan & Ashby, 1992). However, whatever its nature, the mechanism underlying the enumeration of small sets up to four is a good example of a fast numerical process delivering responses that are not necessarily retrieved from memory (but see Logan & Zbrodoff, 2003, or Mandler & Shebo, 1982, for a divergent point of view).

4.3. Procedure limitations

The most striking finding of our study is probably that the size effect in small additions does not exhibit a monotonic increase with the magnitude of the operands, affecting primarily what we called the *very small* problems, but not the *medium small* or the $n + 1$ problems. What distinguishes these latter types of problems from the *very small* additions is the fact that they involve operands larger than four. If the size effect in *very small* additions reflects, as we assume, the use of the compiled procedure that we described above, this means that this procedure cannot operate on operands larger than four. Interestingly, the limit to four is ubiquitous in numerical cognition and more generally in cognitive processes. In the domain of numerical cognition, this limit has been observed in animals (Boysen & Berntson, 1989; Brannon & Terrace, 1998), in human infants when discriminating (Antell & Keating, 1983; Starkey & Cooper, 1980; Strauss & Curtis, 1981) or comparing sets of objects (Feigenson, Carey, & Hauser, 2002), in toddlers and young children when anticipating the result of numerical transformations (Hughes, 1986; Starkey, 1992), as well as in adults when subitizing collections (Mandler & Shebo, 1982; Trick & Pylyshyn, 1994). It has been suggested that this boundary reflects a broader capacity limit of the human cognitive system due to the limitation of the focus of attention to four chunks (Cowan, 2001; Cowan, 2005).

The limitation to four of the number of elements that can be individuated within a single representation and made readily available for further treatment could account for the limitations of the compiled procedure we described above. We have suggested that such a procedure would scan some analogical representation of each of the operands, recursively accessing knowledge related with the number chain and tagging each representational element with the next numerical value retrieved from long-term memory (Fig. 9). Such a non-verbal counting has been evoked by Gallistel and Gelman (1992), Gallistel and Gelman (2000; see also Gelman & Tucker, 1975), with the difference that these authors assume a scalar variability in the magnitudes that would represent discrete (countable) quantities in the same way as continuous uncountable quantities (see below for a discussion of this view). Instead, the representations we hypothesize are closer to an object file in which each object is individuated (Feigenson et al., 2002; Simon, 1997). It is known that the use of object-file representations yields a set-size signature with success at representing sets up to four objects and failure with larger numbers (Kahneman, Treisman, & Gibbs, 1992; Pylyshyn, 1989; Pylyshyn, 1994). The involvement of this type of representation in *very small* problems would explain why their solving was affected by a strong size-effect, whereas RTs for small problems involving operands larger than four (the *medium small* problems) remained immune to size effect. In the same way as 12-month-old infants are no longer able to reliably compare two sets of objects when one of them involves more than three objects (Feigenson et al., 2002), any operand larger than four could not give rise to the precise analogue representation on which the fast compiled procedure operates, and subjects would be obliged to switch to another strategy. This limitation makes that the fast procedure we assume for *very small* additions differs from the *Min* strategy described by Groen and Parkman (1972). Whereas the *Min* strategy can be used for any pair of operands as long as one of them is sufficiently small, the use of the compiled procedure is restricted to problems in which both operands can be automatically converted into analogical representations (i.e., no larger than four).

It is worth to note that, in line with models of working memory which suppose that only one item can be processed at a time (Barrouillet & Camos, 2015; Oberauer, 2002), our model assumes that the two operands are successively, and not simultaneously, converted into the analogical representation on which the compiled procedure operates. This explains why problems with operands that do not exceed 4 can be processed by the compiled procedure even when their sum exceeds 4. Indeed, it could be argued that encoding the first operand of the problem $4 + 3$, for example, would exhaust working memory capacity, preventing the conversion of the second operand and blocking the compiled procedure. However, we conceive working memory as a dynamic system the content of which is continuously updated. Thus, once the first operand has been converted into its analogical representation and processed, this representation is no longer needed and attention can move to the second operand for encoding purpose with the entire working memory capacity (limited to 4) available anew (see Fig. 9).

The involvement of working memory on the construction and processing of these analogical representations explains why the functioning of the compiled procedure is constrained by working memory capacity, individuals with lower capacity exhibiting steeper size-related slopes. It could be surprising that a compiled and automated procedure is constrained by working memory capacity. However, it has been shown that the speed of even very elementary cognitive processes such as retrieving overlearned information (e.g., reading Arabic digits) or subitizing small sets of objects is constrained by working memory capacity (Barrouillet et al., 2008). The compiled procedure we hypothesize involves encoding numbers, scanning representations, and the recursive access to long-term knowledge, all processes that are faster in high-capacity individuals, the concatenation of small differences on each of these processing steps underpinning the individual differences reported on Section 3.4 above.

4.4. Nature of strategies and strategy choice

A pending question concerns the nature of the strategies used for solving the other small problems like *medium small* or $n + 1$ problems. The absence of any identifiable size-effect in these two categories of problems led us to assume that *medium small* problems are solved through direct retrieval of their answer from memory, whereas $n + 1$ problems would involve a one-greater rule. A striking and probably meaningful phenomenon is that these two types of problems have mean RTs corresponding to the slowest *very small* problems with the same structure. More precisely, in the subset of the 51 frequent retrievers, $n + 1$ problems (with $n > 4$) have a mean RT of 817 ms which is close to the mean RT of the slowest *very small* problems involving 1 (i.e., 1 + 4 and 4 + 1, mean RT of 796 ms). In the same subset of participants, the mean RT for *medium small* problems (928 ms) was very close to the mean RT of the slowest *very small* problems (i.e., 3 + 4 and 4 + 3, mean RT of 912 ms). It is as though solving times for small problems increased linearly with the size of the operands up to four, then plateauing (see Fig. 5). Though we have no explanation for this phenomenon, some leads can be suggested.

A first possibility is to assume some strict order in the strategies successively implemented by problem solvers when encountering a problem, as Siegler and Shrager (1984) did in their distribution of associations model, but in a different order. Siegler and Shrager (1984) proposed that retrieval was first attempted, a representation of the problem being elaborated when retrieval fails. If no answer can be read out from this representation, problem solvers would use some back up algorithmic strategy. It could be imagined instead that the first attempt consists of constructing the analogue representation on which the *next-token-next-value* production is applied. This type of encoding could be privileged when numbers are encoded for transformation purpose, as it is the case in arithmetic operations (Thevenot & Barrouillet, 2006). When this representation cannot be constructed because at least one of the operands is larger than four, the decision cycle corresponding to the fast procedure does not deliver any answer, and the subject would try to retrieve this answer from associations stored in long-term memory. This would explain why problems solved by direct retrieval such as the *medium small* problems do not result in faster responses than the *very small* problems solved using the prioritized fast procedure.

Another possibility, inspired by Logan's (1988) race model, could be that algorithmic procedures race against the direct retrieval strategy, algorithms winning the race as long as they reach their goal before retrieval. This would be the case for *very small* additions that can be solved by the compiled procedure implemented in the analogue representations evoked above. However, for problems involving operands larger than four and for which the compiled procedure cannot be used, retrieval would at the end be the fastest strategy because the algorithmic procedures used for solving these problems in the first place cannot reach a sufficient level of automation. Overall, the surprising size-effect pattern displayed in Fig. 5 with a strong size-related increase of RTs followed by a plateau remains difficult to explain within the traditional explanatory frameworks of cognitive arithmetic.

4.5. Alternative accounts

We have assumed that the strong size-effect affecting *very small* additions reflects a procedural solving. We discarded above several accounts of this effect revolving around the idea that a unique retrieval process could vary in latency from problem to problem due to history of acquisition, frequency of exposure, or susceptibility to interference. None of these factors can reasonably produce the linear trends we observed with steep slopes of more than 45 ms per increment. However, the size effect affecting small additive problems has received other explanations that aimed at preserving the retrieval hypothesis. This is the case of the hypotheses of a tabular search, of a mediation of retrieval via mappings from written numbers to preverbal magnitudes, and of a sporadic use of slow counting strategies when retrieval fails.

Let us begin by the tabular search hypothesis. It is often assumed in the literature that problem-answer associations do not remain unorganized in long-term memory, but take place within some square table with entry nodes for the digits 0–9 on two adjacent sides (Ashcraft & Battaglia, 1978). Response times would correspond to the time required to search the point of intersection corresponding to the operands. Searching this kind of table could account for a linear increase of RTs with operands magnitude. However, we saw that *medium small* additions did not exhibit the size effect that this hypothesis would predict.

The second hypothesis, closer to our model of an automated compiled procedure, was proposed by Gallistel and Gelman (1992) within their model of pre-verbal counting. Following a hypothesis from Restle (1970), Gallistel and Gelman assume that number facts are retrieved by mapping numerical values to magnitudes corresponding to positions on the number line. First, each addend would be mapped onto magnitude representations, these magnitudes being preverbally added to obtain a new magnitude corresponding to the sum of the addends. Mapping this resulting magnitude back to the verbal domain would provide the answer. The explanation of the size effect runs as follows. It is assumed that numerosities are represented by magnitudes of an increasing variability that obeys Weber's law: the larger the numerosity, the greater the variability of the preverbal representation, a phenomenon referred to as "scalar variability". This increasing variability makes that the greater the numerosities represented, the higher the variability of the corresponding magnitudes, and the more difficult their accurate mapping onto the corresponding numerical value. Thus, the larger the sum, the higher the scalar variability of the represented magnitude, and the longer the mapping process must wait in order to obtain a mapping of acceptable reliability. In other words, the size effect in additions would result from a speed-accuracy trade-off in mapping back added magnitudes to the corresponding numerosity (i.e., the answer). Although this account relies on a pre-verbal counting hypothesis akin to the process we described above, the predictions of this model are at odds with our data. Indeed, contrary to the object file hypothesis, the models based on scalar variability cannot predict any discontinuity in magnitude effects. Thus, it predicts a monotonic increase in RTs with the size of the problem whereas we observed a non-monotonic size effect with a plateau when operands are larger than four.

Another way of explaining how retrieval-based solving could produce a size effect is to assume, as Groen and Parkman (1972) did, that the size-related slope of RTs is due to an artefact of averaging. They explained the size-related slope of some tens of milliseconds observed in adults by assuming that they solve additions by fast retrieval, but occasionally revert with some probability p to a childish way of solving additions and use a slow counting procedure such as the *Min* strategy. A uniform probability p over problems would result in the linear trend we observed. It is fairly possible that a retrieval process fails for some undetermined reason in a small proportion of trials. However, this hypothesis cannot account for our results. First, like the tabular search hypotheses, the sporadic slow counting hypothesis predicts a monotonic size effect that we did not observe. There is indeed no reason to suppose that retrieval would sometimes fail for *very small* additions, but never for *medium small* additions that did not exhibit any sizeable slope related with magnitude. Moreover, it is difficult to explain within this account that *very small* additions exhibited a slope significantly steeper than ties that are most probably solved through retrieval without the additional ad hoc assumption that the probability *p* of reverting to slow counting is far lower for ties than *very small* additions. As a consequence, it seems that it is impossible to reconcile our findings with the hypothesis that *very small* additions are solved through retrieval of their answer from memory.

5. Conclusions

Chronometric analyses contradict the hypothesis that educated adults solve the smallest and most frequent additions by retrieving their answer from long-term memory. This finding does not undermine the hypothesis that problem-answer associations are stored in long-term memory and we have seen that among small additions, many problems are probably solved by retrieval. However, it challenges the received view that retrieving answers from memory is necessarily faster than any other process in problem solving, and as such the hallmark of expertise. As Baroody (1983) surmised many years ago, automated compiled procedures can be so fast in their execution that people remain unaware of their process, mistaking them with a retrieval from memory. This speed of implementation along with the lack of awareness creates a misleading convergence of indices that led psychologists to trust on participants' verbal reports and endorse the retrieval hypothesis without further enquiry. However, a detailed analysis of RTs based on a large set of data reveals that trials reported as retrieved subsume a variety of strategies in which genuine retrieval from long-term memory coexists with fast procedures resulting in strong size effects and probably rules such as that used for $n + 1$ problems. Thus, the prevailing description of development as a progressive shift from algorithmic computing to direct retrieval is probably an oversimplification.

References

Anderson, J. R. (1974). *Human associative memory*. New York: John Wiley & Sons.
Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Antell, S. E., & Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child Development, 54*(3), 695–701.
Ashcraft, M. H. (1982). The development of mental arithmetic: A chronometric approach. *Developmental Review, 2*, 213–236.
Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition, 44*, 75–106.
Ashcraft, M. H., & Battaglia, J. (1978). Cognitive arithmetic: Evidence for retrieval and decision processes in mental addition. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 527–538.
Ashcraft, M. H. (1987). Children's knowledge of simple arithmetic: A developmental model and simulation. In C. J. Brainerd, R. Kail, & J. Bisanz (Eds.), *Formal methods in developmental research* (pp. 302–338). New York: Springer-Verlag.
Ashcraft, M. H., & Fierman, B. A. (1982). Mental addition in third, fourth, and sixth graders. *Journal of Experimental Child Psychology, 33*, 216–234.
Ashcraft, M. H., & Guillaume, M. M. (2009). Mathematical cognition and the problem size effect. In B. Ross (Ed.). *The psychology of learning and motivation* (Vol. 51, pp. 121–151). Burlington: Academic Press.
Ashcraft, M. H., & Stazyk, E. H. (1981). Mental addition: A set of three verification models. *Memory and Cognition, 9*, 185–196.
Balakrishnan, J. D., & Ashby, F. G. (1992). Subitizing: Magical numbers or mere superstition? *Psychological Research, 54*, 80–90.
Baroody, A. J. (1983). The development of procedural knowledge: An alternative explanation for chronometric trends of mental arithmetic. *Developmental Review, 3*, 225–230.
Baroody, A. J. (1984). A reexamination of mental arithmetic models and data: A reply to Ashcraft. *Developmental Review, 4*, 148–156.
Baroody, A. J. (1994). An evaluation of evidence supporting fact-retrieval models. *Learning and Individual Differences, 6*, 1–36.
Barrouillet, P., & Camos, V. (2015). *Working memory: Loss and reconstruction*. Hove, UK: Psychology Press.
Barrouillet, P., & Fayol, M. (1998). From algorithmic computing to direct retrieval. Evidence from number- and alphabetic–arithmetic in children and adults. *Memory & Cognition, 26*, 355–368.
Barrouillet, P., Lépine, R., & Camos, V. (2008). Is the influence of working memory capacity on high level cognition mediated by complexity or resource-dependent elementary processes? *Psychological Bulletin and Review, 15*, 528–534.
Barrouillet, P., & Thevenot, C. (2013). On the problem-size effect in small additions: Can we really discard any counting-based account? *Cognition, 128*(1), 35–44.
Boysen, S. T., & Berntson, G. G. (1989). Numerical competence in a chimpanzee (Pan troglodytes). *Journal of Comparative Psychology, 103*, 23–31.
Brannon, E. M., & Terrace, H. S. (1998). Ordering of the Numerosities 1 to 9 by monkeys. *Science, 282*(5389), 746–749.
Campbell, J. I. D. (1987a). Network interference and mental multiplication. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(1), 109–123.
Campbell, J. I. D. (1987b). The role of associative interference in learning and retrieving arithmetic facts. In J. A. Sloboda & D. Rogers (Eds.). *Cognitive processes in mathematics. Keele cognition seminars* (Vol. 1, pp. 107–122). New York: Clarendon Press/Oxford University Press.
Campbell, J. I. D., Chen, Y., & Maslany, A. J. (2013). Retrieval-induced forgetting of arithmetic facts across cultures. *Journal of Cognitive Psychology, 25*(6), 759–773.
Campbell, J. I. D., & Gunter, R. (2002). Calculation, culture, and the repeated operand effect. *Cognition, 86*, 71–96.
Campbell, J. I. D., & Timm, J. C. (2000). Adults' strategy choices for simple addition: Effects of retrieval interference. *Psychonomic Bulletin & Review, 7*(4), 692–699.
Carpenter, T., & Moser, J. (1983). The acquisition of addition and subtraction concepts. In R. Lesh & M. Landau (Eds.), *Acquisition of mathematics: Concepts and processes* (pp. 7–44). New York: Academic Press.
Collins, A. M., & Quillian, M. R. (1972). How to make a language user. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 309–351). New York: Academic Press.
Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–185.
Cowan, N. (2005). *Working memory capacity*. Hove: Psychology Press.
Fayol, M., & Thevenot, C. (2012). The use of procedural knowledge in simple addition and subtraction problems. *Cognition, 123*, 392–403.
Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying children's choice of more: Object files versus analog magnitudes. *Psychological Science, 13*(2), 150–156.
Ferrand, L., Brysbaert, M., Keuleers, E., New, B., Bonin, P., Meot, A., ... Pallier, C. (2011). Comparing word processing times in naming, lexical decision, and progressive demasking: Evidence from chronolex. *Frontiers in Psychology, 2*.
Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*(1–2), 43–74.
Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences, 4*(2), 59–65.
Geary, D. C., & Brown, S. C. (1991). Cognitive addition: Strategy choice and speed-of-processing differences in gifted, normal, and mathematically disabled children. *Developmental Psychology, 27*, 398–406.
Geary, D. C., & Burlingham-Dubree, M. (1989). External validation of the strategy choice model for addition. *Journal of Experimental Child Psychology, 47*, 175–192.
Gelman, R., & Tucker, M. F. (1975). Further investigations of the young child's conception of number. *Child Development, 46*, 167–175.
Girelli, L., Lucangeli, D., & Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *Journal of Experimental Child Psychology, 76*(2), 104–122.
Groen, G. J., & Parkman, J. M. (1972). A chronometric analysis of simple addition. *Psychological Review, 79*, 329–343.
Hamann, M. S., & Ashcraft, M. H. (1986). Textbook presentations of the basic addition facts. *Cognition and Instruction, 3*, 173–192.
Hughes, M. (1986). *Children and number*. Oxford: Blackwell Publishers.
Imbo, I., & Vandierendonck, A. (2007). The development of strategy use in elementary school children: Working memory and individual differences. *Journal of Experimental Child Psychology, 96*, 284–309.
Imbo, I., & Vandierendonck, A. (2008). Effects of problem size, operation, and working memory span on simple-arithmetic strategies: Differences between children and adults? *Psychological Research, 72*(3), 331–346.
Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*, 175–219.
Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189–217.
Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology, 62*(4), 498–525.
Kirk, E. P., & Ashcraft, M. H. (2001). Telling stories: The perils and promise of using verbal reports to study math strategies. *Journal of Experimental Psychology, 27*, 157–175.
Klahr, D. (1973). A production system for counting, subitizing, and adding. In W. G. Chase (Ed.), *Visual information processing* (pp. 527–546). New York: Academic press.
Klahr, D., & Wallace, J. G. (1976). *Cognitive development: An information processing view*. Hillsdale, NJ: Erlbaum.
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, 22*, 216–230.
Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492–527.
Logan, G. D., & Zbrodoff, N. J. (2003). Subitizing and similarity: Toward a pattern-matching theory of enumeration. *Psychonomic Bulletin & Review, 10*, 676–682.

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology, 111*, 1–22.
Miller, K. F., Perlmutter, M., & Keating, D. (1984). Cognitive arithmetic: Comparison of operations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 46–60.
Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 28*, 411–421.
Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition, 32*, 65–97.
Pylyshyn, Z. W. (1994). Some primitive mechanisms of spatial attention. *Cognition, 50*, 363–384.
Qin, S., Cho, S., Chen, T., Rosenberg-Lee, M., Geary, D. C., & Menon, V. (2014). Hippocampal-neocortical functional reorganization underlies children's cognitive development. *Nature Neuroscience, 17*, 1263–1269.
Reder, L. M. (1988). Strategic control of retrieval strategies. *The Psychology of Learning and Motivation, 22*, 227–259.
Restle, F. (1970). Speed of adding and comparing numbers. *Journal of Experimental Psychology, 83*, 274–278.
Roussel, J. L., Fayol, M., & Barrouillet, P. (2002). Procedural vs. direct retrieval strategies in arithmetic: A comparison between additive and multiplicative problem solving. *European Journal of Cognitive Psychology, 14*, 61–104.
Rubinsten, O., Henik, A., Berger, A., & Shahar-Shalev, S. (2002). The development of internal representations of magnitude and their association with Arabic numerals. *Journal of Experimental Child Psychology, 81*, 74–92.
Siegler, R. S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229–293). Hillsdale, NJ: Erlbaum.
Simon, T. (1997). Reconceptualizing the origins of number knowledge: A non-numerical account. *Cognitive development, 12*, 349–372.
Sokol, S. M., McCloskey, M., Cohen, M. J., & Aliminosa, D. (1991). Cognitive representations and processes in arithmetic: Inferences from the performance of brain-damaged subjects. *Journal of Experimental Psychology, 17*, 355–376.
Starkey, P. (1992). The early development of numerical reasoning. *Cognition, 43*, 93–126.
Starkey, P., & Cooper, R. G. (1980). Perception of numbers by human infants. *Science, 2010*(4473), 1033–1035.
Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153*, 652–654.
Strauss, M. S., & Curtis, L. E. (1981). Infant perception of numerocity. *Child Development, 52*, 1146–1152.
Thevenot, C., & Barrouillet, P. (2006). Encoding numbers: Behavioral evidence for processing-specific representations. *Memory & Cognition, 34*(4), 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, 54*(2), 599–611.
Thorndike, E. L. (1922). *The psychology of arithmetic*. New York: The Macmillan Company.
Trick, L. M. (1992). Subitizing and counting: Preattentive and attentive processing in visual enumeration. *Dissertation Abstracts International, 52*(11-B), 6117.
Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review, 101*(1), 80–102.
Widaman, K. F., Geary, D. C., Cormier, P., & Little, T. (1989). A componential model for mental addition. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15*(5), 898–919.
Zbrodoff, N. J. (1995). Why is 9 + 7 harder than 2 + 3? Strength and interference as explanation of the problem size effect. *Memory and Cognition, 23*, 689–700.
Zbrodoff, N. J., & Logan, G. D. (2005). What everyone finds: The problem-size effect. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 331–345). New York and Hove: Psychology press.