On the problem-size effect in small additions: Can we really discard any counting-based account?

Barrouillet, Pierre Noël; Thevenot, Catherine

How to cite

BARROUILLET, Pierre Noël, THEVENOT, Catherine. On the problem-size effect in small additions: Can we really discard any counting-based account? In: Cognition, 2013, vol. 128, n° 1, p. 35–44. doi: 10.1016/j.cognition.2013.02.018

This publication URL: https://archive-ouverte.unige.ch/unige:88203
Publication DOI: 10.1016/j.cognition.2013.02.018

Cognition 128 (2013) 35–44

Contents lists available at SciVerse ScienceDirect

Cognition

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

On the problem-size effect in small additions: Can we really discard any counting-based account?

Pierre Barrouillet *, Catherine Thevenot *

Université de Genève, Faculté de Psychologie et des Sciences de l'Education, 40 bd du pont d'Arve, 1205 Genève, Switzerland

ARTICLE INFO

Article history:
Received 13 July 2012
Revised 12 November 2012
Accepted 27 February 2013
Available online 9 April 2013

Keywords:
Numerical cognition
Mental addition
Strategies
Arithmetic
Working-memory

ABSTRACT

The problem-size effect in simple additions, that is the increase in response times (RTs) and error rates with the size of the operands, is one of the most robust effects in cognitive arithmetic. Current accounts focus on factors that could affect speed of retrieval of the answers from long-term memory such as the occurrence of interference in a memory network or the strength of memory traces that would differ from problem to problem. The present study analyses chronometric data from a sample of 91 adults solving very small additions (operands from 1 to 4) that are generally considered as being solved by retrieval. The results reveal a monotonic linear increase in RTs with the magnitude of both operands. This pattern is at odds with the retrieval-based accounts of the problem-size effect and challenges the well-established view that small additions are solved through retrieval of the answer from long-term memory. Our results are more compatible with the hypothesis that even very small additions are solved using compacted fast procedures that scroll an ordered representation such as a number line or a verbal number sequence. This interpretation is corroborated by the analysis of individual differences.

1. Introduction

The problem-size effect, that is the increase in response times (RTs) and error rates with the size of the operands of additions and multiplications, is a benchmark of mental arithmetic (Zbrodoff & Logan, 2005). Concerning additions, on which this article focuses, four decades of investigation have not permitted to reach a consensus on the sources of this effect. Though the selection of nonretrieval and slower procedures on larger problems (i.e., additions with a sum larger than 10) has been advocated to account for longer solution times with larger problems (LeFevre, Sadesky, & Bisanz, 1996), a problem size effect is also observed for additions with a sum up to 10 (hereafter, *small* additions) or when analyses are restricted to those trials in which participants report to have solved additions by directly retrieving the answer from memory. In these latter cases, the leading hypotheses focus on factors that could affect speed of retrieval such as the occurrence of interference in a memory network (Zbrodoff & Logan, 2005) or the strength of memory traces that would differ from problem to problem (Ashcraft & Guillaume, 2009). In the present study, we concentrate on very small additions (augends and addends, i.e., the first and second operands respectively, from 1 to 4) for which there is a consensus on the idea that they are solved through retrieval. Contrasting with this received view, the analysis of RTs from a large sample of adult participants reveals a clear problem-size effect that does not fit very well with any of the retrieval-based accounts.

In a seminal study, Groen and Parkman (1972) presented first graders with simple addition problems (the largest problem was 5 + 4) and asked them to select an answer from 0 to 9 by pressing labeled buttons. The results are well known. The best predictor of RTs proved to be the smaller of the two addends. This phenomenon pointed

* Corresponding authors. Tel.: +41 223799251.
E-mail addresses: Pierre.Barrouillet@unige.ch (P. Barrouillet), Catherine.Thevenot@unige.ch (C. Thevenot).

0010-0277/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cognition.2013.02.018

towards a *Min* model assuming that a counting procedure starts from the larger addend and then counts on by ones for the value of the smaller addend (e.g., for 2 + 4, counting 4, 5, 6). Regression analyses revealed a slope of 410 ms in RTs, suggesting that each increment in the counting procedure took more than 400 ms in young children, hence producing a large problem-size effect. An exception to this trend was noted for tie problems (e.g., 3 + 3) that were solved faster and seemed to not exhibit any size effect. Consequently, Groen and Parkman suggested that the answers of tie problems were retrieved from long-term memory, an idea that is now universally admitted. Investigations in adults led to a different picture with a slope for the *min* of only 20 ms that was judged an implausibly fast rate for a counting procedure. Groen and Parkman suggested that instead of counting, adult directly retrieve the answer of small additions from memory. The small problem-size effect would result from the sporadic use of counting strategies in the rare trials (about 5%) on which direct access failed.

The idea that the problem-size effect in adults is due to the sporadic use of counting strategies was subsequently criticized by Ashcraft (Ashcraft & Battaglia, 1978; Ashcraft & Stazyk, 1981) who noted that, when tie problems are excluded, the best predictor of RTs in regression analyzes was the square of the sum of the operands, whereas any counting model would predict linear and not exponential increases in RTs with operands size. Consequently, the idea was put forward that simple additions were solved through the search of 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 process would slow down as the search progresses, explaining the non-linear problem-size effect and that the best predictor is the square of the sum rather than the sum itself.

Subsequently, other characteristics of the memory network were advocated to account for the problem-size effect. Siegler and Shrager (1984) suggested that the direct retrieval of the answer cued by the operands can also activate competing associations resulting from the obligatory encoding of errors occurring earlier in development when additions are mainly solved through counting procedures. Because large problems involve more processing steps when solved through counting and are consequently more error prone, their operands would be associated with more competing responses. This increased number of competitors would result in slower retrievals. The learning history of number facts could also have another influence on the memory network. Hamann and Ashcraft (1986) observed that the frequency with which addition problems appear in elementary school textbooks decreases as the size of the operands increases. More often practiced problems would lead to stronger associations with their answer. Both the associations of large problems with more errors and their lower frequency led Ashcraft and Guillaume (2009) to assume that they are less strongly represented in memory, hence their slower retrieval.

Akin to this memory strength model is the network interference model proposed by Zbrodoff and Logan (2005) who emphasize how the memory network is susceptible to interference with 100 pair-wise associations between only 10 digits. Using alphabetic arithmetic (e.g., A + 3 = ?), Zbrodoff (1995) was able to independently manipulate level of practice of a given problem and similarity among studied problems by contrasting overlapping items prone to interference (e.g., M + 2, M + 3, M + 4) and non-overlapping items (A + 2, D + 3, H + 4). The results revealed that it is the combination of differential practice and interference that makes the problem-size effect appear, and that this effect is not due to counting but to retrieval. From these results and studies on error priming, Zbrodoff and Logan claimed that the interference model "remains the best explanation of memory-based arithmetic performance" (p. 338).

Standing in opposition to these conceptions, LeFevre et al. (1996) proposed that, like children, adults use both retrieval and nonretrieval procedures to solve additions. Using self-report of procedures for solving simple additions (operands from 0 to 9), they observed that retrieval was frequent for small additions (more than 80%) but far rarer for large additions when ties are excluded (only 47%). Thus, they established that the problem-size effect is mainly due to the use of slower nonretrieval strategies on problems with a sum larger than 10. Interestingly, the authors observed that latencies on retrieval trials still increased with problem size, but this effect was substantially reduced when compared with the effect on all the trials and it was not particularly systematic. In accounting for this problem-size effect in retrieval trials, LeFevre et al. favored the idea that retrieval latencies could reflect acquisition history. Problems that were more often solved through nonretrieval procedures in the course of development would lead to flatter distribution of associations between problems and answers resulting from counting errors, as suggested by Siegler and Shrager (1984), and to relatively strengthened nonretrieval procedures. These two factors would concur to produce longer retrieval latencies when the answer is retrieved. Importantly, LeFevre et al. noted that "according to this view, the relation between indices of problem size and latency is likely to be epiphenomenal" (p. 227).

In summary, modern accounts of the problem-size effect on small additions converge on two ideas. The first is that the effect results from some structural or functional characteristics of a process of retrieval from a memory network that stores associations. The second is that, as emphasized by Ashcraft (1992, p. 80), "the term *problem-size effect* is now generally considered as a misnomer" because the effect would not be due to the size of the problem per se. Indeed, for both the memory strength and the network interference models, as well as for LeFevre et al.'s (1996) account of the problem-size effect on retrieval trials, the structural characteristics of the problem (i.e., magnitude of the *min*, the sum, or its square) are only coincidentally predictive of RTs in virtue of their relationship to a more central variable that is problem difficulty, which determines both the variability in memory strength and the amount of interference within an organized long-term memory representation of fact knowledge.

Despite a large consensus on the fact that the problem-size effect in small additions results from variations in a retrieval process and not from a counting procedure, some uncertainty remains about the real nature of the processes underpinning addition problem solving. First, the main argument on which the rejection of any counting model was initially based is the non-linear (exponential) increase in RTs with operand size observed by Ashcraft and Battaglia (1978). However, these results were based on a very small sample of 12 participants who verified each of the 100 one-digit additions only once with a correct sum and once with an incorrect sum. The square of the sum was the best predictor of the true problems, but when the 200 problems were considered (including false problems), their RTs correlated at .512 with the correct sum and at .498 with its square. The fact that the two predictors were almost collinear ($r = .966$) renders their comparison in linear regressions poorly informative. Ashcraft and Stazyk (1981, Experiment 1) also studied a small sample (20 participants) using the same procedure. The correlation matrix for RTs on true non-tie problems and different predictors revealed that *min*, sum, and sum^2 provided approximately the same fit (r values of .777, .737, and .775 respectively). In the same way, LeFevre et al. (1996) reported a not particularly systematic problem-size effect on retrieval trials, but their analyses were based on the responses of 16 adults who solved each of the 100 simple additions only once. Thus, empirical evidence of a non-linear or nonsystematic relationship between RTs and magnitude of the operands is less compelling than usually assumed.

Second, some recent studies have called into doubt the idea that small additions are necessarily solved through direct retrieval. Fayol and Thevenot (2012) observed that when the sign of the operation appears 150 ms before the operands, additions and subtractions are solved faster than when the sign and the operands are displayed simultaneously on screen (see also Roussel, Fayol, & Barrouillet, 2002). This priming effect, which was not observed for multiplication problems, was interpreted by the authors as evidence that additions and subtractions are solved using compacted counting procedure that can be activated by the presentation of the sign. The fact that something was activated whereas the operands are not yet presented suggests that the activated knowledge concerns a class of problems rather than a given problem. Such general knowledge cannot correspond to a declarative chunk storing a specific operand-answer association, but rather to a procedure that can be implemented by a variety of operands. Interestingly, this priming effect was observed for small, and even very small, additions (e.g., 2 + 3), but not for ties. If it turned out, as Fayol and Thevenot claim, that small additions are solved through reconstructive strategies such as compacted counting procedures, retrieval-based models of the problem-size effect should be revisited. As the small samples used in some studies on the problem-size effect suggest, and contrary to Ashcraft and Guillaume (2009) who stated that a poor way to resolve the question is to continue to pursue the research using current methodologies such as chronometric measures, we think that reliable sets of data from large samples are needed to precisely analyze the relationship between RTs and the magnitude of operands.

In the following, we present chronometric data from a large pool of 91 participants who were asked to solve very small additions involving operands from 1 to 4, each participant performing 6 trials on each of the 16 additions. This restricted set of additions was initially chosen for a purpose totally different from that which guides the present analysis. It was used to study the impact of individual differences in working memory capacity on elementary processes including also subitizing and reading digits as well as on more complex processes such as counting large arrays of dots (Barrouillet, Lépine, & Camos, 2008). Analyses in this previous publication did not go beyond the mere observation that there was a problem-size effect for additions that was stronger in low- than in high-span participants. No systematic comparison of the different additive problems was run but, as we will see, the data are of high interest for the question of the problem-size effect. Moreover, these data are analyzed as a function of working memory capacities. It has been argued that high working memory capacities would facilitate both the association between operands and answers in long term memory and their fast and reliable retrieval (Geary, Bow-Thomas, Liu, & Siegler, 1996; Thevenot, Barrouillet, & Fayol, 2001). Indeed, associations between operands and answers in memory are assumed to result from the repeated practice of algorithmic procedures during childhood (Siegler & Shrager, 1984). However, lower working memory capacities could involve slower counting procedures resulting in longer delays between problem encoding and production of the answer and more frequent errors. Increased delays and frequent errors would reduce the probability of encoding associations between problems and answers and compromise their subsequent retrieval (Thevenot et al., 2001). Moreover, several models conceive working memory capacity as the available amount of resource needed for activating and retrieving items of knowledge from long-term memory (Barrouillet, Bernardin, & Camos, 2004; Cowan, 1999; Lovett, Reder, & Lebière, 1999). Thus, lower working memory capacities should be associated with rarer retrievals, something that has already been observed in children (Barrouillet & Lépine, 2005; Barrouillet, Mignon, & Thevenot, 2008).

2. Method

2.1. Participants

Ninety-two undergraduate French-speaking students from the Université de Bourgogne (76 females, Mean age: 20 years, $SD = 3.35$ years) received course credit to participate.

2.2. Material and procedure

The addition task analyzed here was performed along with other numerical and two working memory tasks, a reading span task and an alphabet recoding task (see Barrouillet, Lépine, et al., 2008, for details). For the reading span task, three blocks of 20 sentences each were printed in a booklet with one sentence per page. Each block contained

5 series from 2 to 6 sentences each presented in ascending order. Participants were asked to successively read out loud the sentences and to remember their final word. When the participant failed to correctly recall a given series, the presentation of the block was interrupted and the next block was presented with the same procedure. The span was the total number of words in the experimental series correctly recalled (from 0 to 60). The alphabet recoding is a task in which participants perform additions and subtractions in the alphabet considered as a numeric chain (e.g., D − 2 = B; Woltz, 1988). Twenty-four random three-letter sets were created and randomly allocated to six operations (+1, +2, +3, −1, −2, −3), each operation receiving four sets. The three letters appeared successively on screen for 1 s, followed by the operation for 2 s. When a question mark appeared, the participants had to perform the mental operations and to calculate the entire response before writing it down in a notebook. The score was the number of sets correctly converted (from 0 to 24). The scores of the two working memory tasks significantly correlated ($r = .43$, $p < .001$). Thus we calculated a compound score by averaging the two z scores.

As far as the addition task was concerned, participants were asked to add two digits presented simultaneously on screen. Each trial was preceded by a ready signal (a star) for 1 s in the center of the screen that was replaced by two digits displayed in two squares of 5.5 cm disposed side by side, with a space of 3.5 cm between them. A voice-key stopped the timer when participants gave their response, the array disappeared from the screen, and the experimenter keyed the response to record accuracy. Participants pressed the mouse button to display the next problem. Six experimental trials for each of the 16 possible pairs of digits from 1 to 4 (resulting in 96 experimental trials) were presented in a random order. The experimental trials were preceded by 16 practice trials in which all the possible pairs were presented once.

3. Results

Error rate on the addition task was very low and varied from 0% to 0.7% across problems (mean = 0.4%) but was poorly related with problem size ($r = .203$ between error rate and the sum). Among the response times (RTs) of the correctly responded trials, those that differed from the mean of the overall sample of participants by more than three standard deviations were considered as outliers and discarded from the analyses. This procedure resulted in discarding 3% of the reaction times on the correct responses. Two additional percents were lost following voice-key failures. One participant was discarded from the RT analyses because he did not have any RT on the addition 4 + 2 after removing errors, outliers, and voice-key failures. Mean RTs for each of the 91 remaining participants on each of the 16 problems were then calculated and used for further analyses (Table 1). We first present the results concerning the entire sample of participants before addressing working memory-related individual differences. In each case, we first investigated the problem-size effect by analyzing through ANOVAs how the size of the two operands affected RTs. The determinants of this size effect were then explored by regression analyses.

Table 1
Mean responses times (*SDs*) for the 16 problems studied in the entire sample.

Augend	Addend			
	1	2	3	4
1	590 (71)	676 (78)	695 (104)	731 (125)
2	673 (96)	643 (80)	713 (116)	768 (145)
3	687 (110)	709 (115)	699 (101)	752 (140)
4	709 (110)	748 (141)	731 (132)	642 (81)

3.1. Overall analyses

First, we performed an ANOVA on the mean RTs for correct responses with the size of the first and second operands as within-subject factors. This analysis revealed a strong problem-size effect with RTs increasing with the size of both the first, $F(3, 270) = 42.27$, $p < .001$, $\eta^2 = .32$, and the second operand, $F(3, 270) = 80.85$, $p < .001$, $\eta^2 = .47$, the two factors significantly interacting, $F(9, 810) = 69.25$, $p < .001$, $\eta^2 = .44$. This interaction was mainly due to shorter responses on tie than non-tie problems (643 ms and 716 ms respectively), $F(1, 90) = 211.94$, $p < .001$, $\eta^2 = .70$, shorter responses that occurred necessarily at different values of the second operand for different values of the first operand (e.g., the tie problem when the first operant is 1 corresponds to a second operand of 1, but the second operand is 2 when the first operand is 2, and so on). Another source of this interaction was an effect of asymmetry with shorter responses on problems with the largest operand as augend than addend (710 ms and 723 ms respectively), $F(1, 90) = 34.08$, $p < .001$, $\eta^2 = .27$.

Second, we investigated the predictive power of structural predictors of the problems on the observed mean RTs (Table 2). Along with the traditional structural predictors (i.e., sizes of the first and second operands, of the minimum addend, of their sum and its square, as well as their

Table 2
Predictors of response times for different sets of problems in the entire sample with correlation coefficients and mean slopes (standard deviations).

Predictor	Problems					
	All[a]		Nonties[b]		Ties[c]	
	r	Slope	*r*	Slope	*r*	Slope
First	.30	12 (14)	.32	8 (17)	.62	21 (16)
Second	.48*	19 (15)	.70**	18 (19)	.62	21 (16)
Minimum	.15	7 (16)	.67*	26 (39)	.62	21 (16)
Sum	.54*	15 (14)	.89**	20 (24)	.62	11 (8)
Sum^2	.43	1 (1)	.86**	2 (2)	.49	1 (1)
Product	.39	4 (4)	.62*	7 (9)	.49	3 (3)
% Retrieval	.14		.51		−.45	
Overlap	.54*		.10		.71	

% Retrieval: percentage of retrieval use reported by the participants in LeFevre et al. (1996). *Overlap*: see text.
[a] Number of problems was 16.
[b] Number of problems was 12.
[c] Number of problems was 4.
* $p < .05$.
** $p < .01$.

product), we also entered in the equation the percentage of retrieval use reported by the participants in LeFevre et al. (1996) for the 16 problems studied here, as well as the number of problems sharing the same sum as the problem at hand (e.g., there is only one problem with a sum of 2 which is 1 + 1, but four problems with a sum of 5 that are 4 + 1, 2 + 3, 3 + 2, and 4 + 1). Problems overlapping by their sum could suffer interference resulting in slower responses due to a fan effect by which retrieval times increase with the number of items of knowledge connected to the same node in the network (Anderson, 1974).

Beginning by non-tie problems, stepwise regression analysis revealed that the sum was the best predictor of RTs, $t = 6.01$, $p < .01$, with no other factor accounting for additional variance. Of course, because linear (i.e., sum) and exponential (i.e., sum^2) predictors were almost collinear here ($r = .99$ for non-tie problems), this analysis cannot be taken as strong evidence against Ashcraft and Battaglia's (1978) hypothesis of a non-linear increase in RTs with operand size. However, at the very least, the present results did not confirm this hypothesis and suggest a linear increase, which is illustrated in Fig. 1. Quite remarkably for a so restricted range of problems, the correlation between RTs and the sum was high ($r = .89$ for nontie problems) with a slope of 20 ms per increment. Excluding tie problems revealed that, as shown in Fig. 1, RTs monotonically increased with both the first and the second operands (with the exception of 4 + 3 that took unexpectedly shorter than 4 + 2). The slopes related with the increase in size of the second operand were of 28 ms, 30 ms, 22 ms, and 11 ms when the first operand was 1, 2, 3, and 4 respectively for a mean of 23 ms, whereas the slopes related with the increase in size of the first operand were of 18 ms, 24 ms, 12 ms and 11 ms when the second operand was 1, 2, 3, and 4 respectively, for a mean of 16 ms. As we will see below, analyses of individual differences will shed light on the small difference between the two mean slopes (i.e., 16 ms for the first operand and 23 ms for the second).

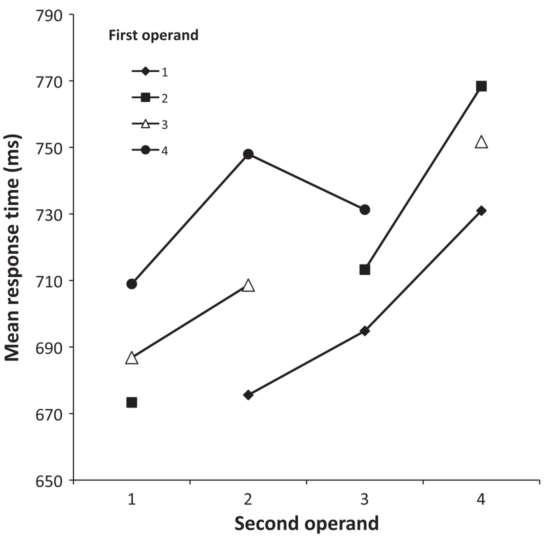

Fig. 1. Mean RTs as a function of the magnitude of the first and second operands for non-tie problems.

Tie problems also exhibited a size effect, with a slope associated with the sum of 11 ms. However, the best predictor of RTs in tie problems was not the sum of the operands or any other structural descriptor, but the amount of overlap between problems, a factor that has no significant effect on RTs for non-tie problems. It can be seen in Table 1 that the problems 1 + 1 and 4 + 4 that had unique sums (i.e., 2 and 8 respectively) involved faster responses than the other tie problems (and the other problems involving these operands). Though the number of tie problems studied did not allow for deeper analyzes, the present results point toward different processes underpinning tie and non-tie problem solving.

Overall, apart from the tie problems, RTs were strongly related with structural characteristics of the problems, and more precisely with the magnitude of the operands, with an increase of about 20 ms in RTs when either the first or the second operand was incremented by one[1]. This is not to say that the problem-size effect affected all the participants. Among the 91 participants, important individual differences were observed in the susceptibility to the size effect that are analyzed in the next section.

3.2. Individual differences in problem-size effect

As we noted above (see Footnote 1), contrary to what was observed with additions, reading digit times in Barrouillet, Lépine, et al. (2008) were immune to size effect with a negligible slope of 2 ms from 2 to 8. On this basis, we decided that a size-related slope that did not exceed 5 ms would correspond to a nil or at least negligible size effect, and we accordingly classified our participants depending on whether they exhibited a size-effect (i.e., a

[1] As suggested by an anonymous reviewer, it could be argued that this problem-size effect resulted from a differential sensitivity of the vocal key to the utterance of the different answers. Though such a factor could have played a role, it does not seem that it was the source of the problem size effect observed. Our data are issued from a larger study in which participants were asked to perform a series of tasks including a reading digit task in which participants were asked to read Arabic digits aloud. If the problem-size effect we observed was due to a differential sensitivity of the voice key, the same size effect would also occur in this reading task. This was not the case. Though reading times varied from one digit to another, the size effect from 2 to 8 (the range of the answers uttered on our study) was negligible (slope of 2 ms) in the entire sample of participants whereas the problem size effect on the same range of additive answers was 15 ms. When considering the subgroup of 44 participants who exhibited a size effect for both operands (see the "individual differences in size effect" section), they exhibited a size-related slope of 24 ms for additions, but of 3 ms for reading digits. In order to go further in discarding the interpretation of our results in terms of differential sensitivity of the voice key, the ANOVA performed on the mean RTs for correct responses with the size of the first and second operands as within-subject factors was replicated on transformed RTs obtained by subtracting from the mean RTs the mean time taken by the entire sample for reading the digit corresponding to the answer. This analysis revealed the same effects as previously observed with a strong problem-size effect with transformed RTs increasing with the size of both the first, $F(3, 270) = 10.59$, $p < .001$, $\eta^2 = .11$, and the second operand, $F(3, 270) = 46.28$, $p < .001$, $\eta^2 = .34$, the two factors still significantly interacting, $F(9, 810) = 18.34$, $p < .001$, $\eta^2 = .17$. Transformed RTs still increased with the size of both operands, as testified by the significant linear trends for the first and second operands, $F(1, 90) = 20.15$, $p < .001$, and $F(1, 90) = 78.11$, $p < .001$, respectively. Thus, the sum of the operands was still the best predictor of these transformed RTs with r values of .637 for all the problems and .774 for non-tie problems.

slope higher than 5 ms) on either the first, the second, or both operands. It turned out that 15 participants out of 91 did not exhibit any size-effect of either the first or second operand, suggesting that they mainly solved additions by a retrieval procedure, whereas 44 participants exhibited size-related slopes of more than 5 ms on both operands. The remaining participants exhibited a size-related slope on either the first or the second operand only (9 and 23 participants respectively). Interestingly, those participants who did not exhibit any size effect were faster in solving additions than those who had a slope for only one of the two operands (mean response times of 629 ms and 694 ms respectively), who were in turn faster than those who exhibited a size-effect for both operands (761 ms), $F(2, 88) = 11.87$, $p < .001$. The three groups differed also in their working memory capacities, with participants that did not exhibit size-effect having higher compound scores than those who had a size effect for at least one operand (mean compound z score of 0.55 and 0.07 respectively), who in turn had higher working memory capacities than participants with addition times affected by the size of both operands (mean z score of -0.21), $F(2, 88) = 5.18$, $p < .01$.

These findings suggest that only a minority of less than 20% of our participants relied on retrieval for solving small additions, resulting in fast responses that were not affected by the size of the operands. All the other participants had solution times that increased with the size of one or both operands, with about 50% of undergraduate students who exhibited a strong size effect for both operands, a group in which the sum accounted for 87% of the variance on solution times that varied from 677 ms for 2 + 1 to 826 ms for 4 + 3. It should be noted that these findings strongly depart from what is usually assumed concerning the predominance of retrieval strategies in solving small additions, and even from the studies that moderated this received conception. For example, LeFevre et al. (1996) reported some use of non-retrieval strategies in small additions, but the rate of retrieval reported by these authors for the non-tie problems studied here was still higher than .80, suggesting a massive use of retrieval that contrasts with our results. Moreover, we have seen that the percentage of retrieval reported by these authors was not a reliable predictor of solution times.

3.3. Individual differences related with working memory capacities

As we previously noted, there are theoretical reasons for assuming that working memory capacity should have an impact on the use of retrieval for solving additions and consequently on the existence and magnitude of a problem-size effect. Interestingly, the compound z score of working memory for the entire group of 91 participants correlated with the mean response times on non-tie ($r = -.36$, $p < .01$) and tie problems ($r = -.33$, $p < .01$). It also correlated with the slopes of the regression of RTs on the sum of the operands for non-tie ($r = -.30$, $p < .01$) and tie problems ($r = -.31$, $p < .01$). The lower the working memory capacities, the slower the responses and the stronger the problem-size effect. To illustrate the differences related to working memory capacities, two extreme groups were contrasted. Participants who obtained positive z scores on both working memory tasks and a mean z greater than 0.67 constituted the high-working memory span group (18 participants: mean alphabet recoding = 19.39, $SD = 2.38$; mean reading span = 20.78, $SD = 4.98$), whereas those who obtained two negative z scores and a mean z lower than -0.67 constituted the low-span group (17 participants: mean alphabet recoding = 5.94, $SD = 3.13$; mean reading span = 7.71, $SD = 2.54$).

First, we performed an ANOVA on the mean RTs with the size of the first and second operands as within-subject factors and the group (high vs. low working memory span) as between-subject factor. High-span were faster than low-span individuals (649 ms and 718 ms respectively), $F(1, 33) = 5.93$, $p < .05$, $\eta^2 = .15$. As already observed in the entire sample, RTs increased with the size of both the first and the second operands with a significant interaction, but more interestingly, the size effect was stronger for low than high span individuals concerning the first operand, $F(3, 99) = 5.12$, $p < .01$, $\eta^2 = .13$, whereas the effect of the second operand did not significantly differ between groups, $F(3, 99) = 2.38$, $p > .05$, $\eta^2 = .07$. Nonetheless, the three factors did not significantly interact, $F(9, 297) = 1.55$, $p > .10$, $\eta^2 = .04$. Moreover, both groups exhibited a strong tie effect (737 ms and 661 ms for non-tie and tie problems respectively in the low-span group, 663 ms and 608 ms respectively for the high-span group), but interestingly this tie effect did not interact with groups, $F(1, 35) = 2.18$, $p > .10$. In the same way, working memory capacity did not interact with the asymmetry effect described above, $F < 1$.

Regression analyses shed light on these phenomena. Beginning with non-tie problems, both groups exhibited a size effect with RTs increasing with the sum of the operands, but this effect was stronger in the low span individuals (mean slope of 26 ms, $SD = 24$ ms) than in the high span individuals (mean slope of 8 ms, $SD = 16$ ms), $t(35) = 2.42$, $p < .05$. As Fig. 2 makes clear, the size of both the first and the second operands affected RTs in the low-span group, whereas RTs in high-span individuals remained unaffected by the size of the first operand and only increased with the second operand. This can be explained by the fact that among the 17 low span individuals, 9 were affected by both operands, 5 by the second operand only and 2 by the first, whereas only one of these participants could be considered as a retriever with no problem-size effect. By contrast, 7 out of the 18 high span individuals did not exhibit any problem-size effect, whereas 7 others were affected by the second operand only, and there were only 4 high span participants who were affected by the magnitude of both operands.

Accordingly, while the sum was the best and a very good predictor of RTs in the low-span group ($r = .94$), the best predictor of RTs in the high-span group was not the sum ($r = .65$), but the size of the second operand ($r = .84$). In the low span group, the mean slope associated with the second operand was 26 ms (more precisely, 27 ms, 29 ms, 35 ms, and 11 ms when the first operand was 1, 2, 3, and 4 respectively), very close to the mean slope associated with the first operand, which was 25 ms (28 ms,

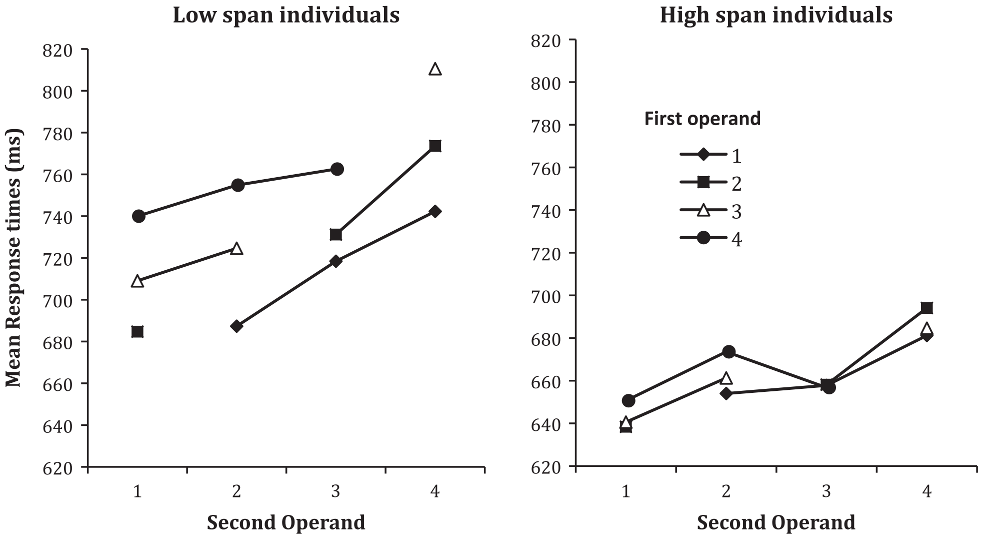

Fig. 2. Mean RTs as a function of the magnitude of the first and second operands for non-tie problems in low-span and high-span individuals.

22 ms, 15 ms, and 34 ms when the second operand was 1, 2, 3, and 4 respectively). By contrast, RTs in high-span individuals were still affected by the magnitude of the second operand (mean slope of 12 ms, with slopes of 14 ms, 17 ms, 14 ms, and 3 ms when the first operand was 1, 2, 3, and 4 respectively) whereas the slope associated with the increase in size of the first operand was practically nil (mean slope of 3 ms). Accordingly, the mean slope associated with the first operand was significantly steeper in low- than in high-span individuals (25 ms compared with 3 ms), $t(35) = 2.94$, $p < .01$, whereas the difference in slopes associated with the second operand did not reach significance (26 ms compared with 12 ms), $t(35) = 1.78$, $p > .05$. In other words, incrementing either the first or the second operand by one resulted in an increase in RT of about 25 ms in the low span group, whereas high-span individuals remained unaffected by the size of the first operand, their RTs increasing by 12 ms when the second operand was incremented by one.

Concerning tie problems, both groups exhibited the same pattern with 1 + 1 and 4 + 4 eliciting the fastest responses (Table 3). Nonetheless, low-span were more affected than high-span individuals by problem size. The sum-related slope was of 11 ms in low-span individuals but only 6 ms in high-span individuals a difference that was highly significant, $t(35) = 3.58$, $p < .01$.

Table 3
Mean responses times (*SDs*) for the 16 problems studied in the high-span and low-span groups.

Augend	Addend			
	1	2	3	4
High span individuals				
1	574 (57)	654 (62)	658 (71)	681 (110)
2	638 (64)	610 (55)	658 (84)	694 (93)
3	641 (71)	661 (101)	646 (81)	685 (108)
4	651 (72)	674 (93)	657 (84)	600 (52)
Low span individuals				
1	602 (59)	687 (80)	718 (119)	742 (109)
2	685 (105)	664 (78)	731 (118)	774 (133)
3	709 (121)	725 (125)	716 (90)	811 (145)
4	740 (110)	755 (113)	763 (120)	661 (66)

4. Discussion

This study explored the problem-size effect in very small additions that are universally considered as solved through direct retrieval of the answer from memory. We used a large sample of participants who solved each problem in six different trials. The results revealed a problem-size effect in non-tie problems that can be unambiguously related with the structural characteristics of the problems, namely the magnitude of the operands. Response times monotonically increased by about 20 ms each time either the augend or the addend was incremented by one. Interestingly, we observed individual differences related to working memory capacities, with higher capacities associated with faster responding and a different pattern of RTs. Whereas individuals with lower working memory capacities exhibited RTs that increased monotonically and in the same extent with both the first and the second operands, individuals with the higher spans remained unaffected by the magnitude of the first operand, but their RTs increased with the size of the second operand. As observed in all studies, tie problems were responded faster. However, their RTs seemed affected by different factors than the non-tie problems. While RTs on these latter problems systematically varied with structural factors, tie problems were primarily affected by interference-related factors such as the degree of overlap between problems. What are the processes underpinning addition problem solving that could produce such a pattern of results? In the following, the capacity of the current models to account for these results is assessed.

One of the main facts arising from this study is that there was no hint in our results for a nonlinear increase in RTs. It should be remembered that Groen and Parkman's seminal (1972) model was rejected on the basis that "no simple counting-based model could account for a nonlinear increase in RT" (Ashcraft & Guillaume, 2009, p. 127). However, as we noted above, linear and nonlinear predictors are almost collinear when considering simple additions, and it is very difficult to assert from regression analyses that a set of data speaks unequivocally for a nonlinear increase. At the very least, our results indicate that, in the restricted range of the problems studied here, a counting-based model cannot a priori be excluded. In the same way, it has been argued that retrieval interference accounts should be preferred to counting and memory search models because these latter models predict monotonic increases in RTs with digit magnitude whereas this increase is actually not strictly monotonic (Zbrodoff & Logan, 2005). Though being not absolutely perfect, the increase in RTs observed here can be considered as reasonably monotonic. A model assuming extra-times of 16 ms and 23 ms for increments by one of the first and second operand respectively provides a very good fit of the mean RTs on the 12 non-tie problems with 84% of variance accounted for. In the subgroup of low-span individuals, this fit increases up to 89%. Thus, assuming that RTs increase with the magnitude of the operands seems to be the simplest account of our results.

Considering that we studied very small additions, such an increase seems difficult to accommodate with either a strength memory model or a network interference model. According to Ashcraft and Guillaume (2009), memory strength would vary depending on acquisition history. Those problems that are difficult when solved through algorithmic strategies during childhood would result in weaker memory traces that take longer to retrieve. Though this hypothesis could account for differences between small and large problems (i.e., with sums larger than 10), it is far less convincing for the very small problems we used. These problems are the easiest that can be found with a minimum addend that does not exceed three, and a gradient of difficulty does not fit very well with the increase in RTs that we have observed. For example, 4 + 1 took longer than 3 + 1, and 4 + 2 longer than 3 + 2 while the difference in difficulty between these problems is far from being evident. In the same way, a network interference model would have difficulties in accounting for these results. According to Zbrodoff and Logan (2005), interference would result from a mix of differential practice and overlap between items. However, the small range of problems we studied probably did not allow for great differences in frequency, and the potential effect of differences in frequency has probably been overestimated in accounting for retrieval times in the domain of mental arithmetic. For example, in the linguistic domain, it has been shown that dramatic differences in word frequency actually result in rather small differences in RTs (e.g., decreasing mean frequency from about 3000 to 60 per million results in an increase in RTs of about 15 ms in a lexical decision task, Ferrand et al., 2011). Moreover, it remains unclear that our problems greatly differed from each other in their amount of overlap when considering operands (all the operands from 1 to 4 were involved in the same number of problems as augend or addend, i.e., 4), and the degree of sum overlap did not correlate significantly with RTs on non-tie problems (see Table 2). Rather, the monotonic and linear increase in RTs that we observed points towards a counting model, or a retrieval model in which some table organized by magnitude would be searched.

Among the possible counting models, it could be imagined, as LeFevre et al. (1996) suggested, that our larger problems (e.g., 3 + 4 or 4 + 2) were more often solved through algorithmic procedures whereas smaller problems would be retrieved. It could also be argued that differences related to working memory capacity would result from the more frequent use of retrieval in high-span individuals. However, two findings contradict this proposal. First, the frequency of reported retrieval use for our problems in LeFevre et al.'s study correlated *positively* with the size of the non-tie problems ($r = .59$) and, consequently, with the RTs we observed (see Table 2), something at odds with the hypothesis of longer RTs resulting from a more frequent use of counting strategies. Second, high-span individuals were not only faster and less affected by the problem-size effect than low-span individuals, but more importantly they did not exhibit the same pattern of solution times, remaining unaffected by the variations in magnitude of the first, but not the second, operand. This cannot be accounted by the mere hypothesis of a higher rate of retrieval strategy in high-span individuals. Another possibility would correspond to the "direct access plus counting" model proposed by Groen and Parkman (1972) who observed in adults a slope of 20 ms very close to our estimate. According to this model, direct access would fail in a small proportion of trials for which a *min* counting process would be used. However, the *min* model does not provide a very good fit of our data (Table 2).

Thus, it seems that we are left with an hypothesis akin to Ashcraft and Battaglia (1978) who suggested that RTs reflect the search in a network representation for additions that would resemble a square table with entry nodes for the digits 0–9 on two adjacent sides. Response times would correspond to the time required to search the point of intersection corresponding to the operands. Ashcraft and Battaglia made additional assumptions such as a stretching of this table on the region of large sums to account for the exponential relationship they observed between RTs and the correct sum, but this additional assumption could be jettisoned if the relationship is linear. However, the psychological plausibility of such a table is questionable. Theoretical accounts favoring direct retrieval from memory as the source of the problem-size effect usually assume that the storage in long-term memory of problem-answer associations is a byproduct of the repeated solving of additive problems through counting strategies. This incidental learning does not follow any particular order, occurring according to the unpredictable encounters with problems, and the process by which the resulting memories would organize themselves as a function of the magnitude of the operands remains unclear.

It is probably premature to propose a final explanation from a single study, but it seems that none of the previous models of the problem-size effect can account for the observed pattern of RT increase with non-tie problems. Thus, the present study permits at least to discard several hypotheses. As we have seen, retrieval-based models advocating either differential memory strength or susceptibility to interference have difficulties in accounting for the monotonic increase in RTs. Moreover, the hypothesis of increased RTs with larger problems due to the more frequent use of counting procedures sporadically used among retrievals is not corroborated by verbal reports of retrieval from LeFevre et al. (1996) study that correlate *positively* with the size of the problems. Finally, the hypothesis of a search in a square table lacks psychological plausibility and is undermined, at least in its received version (Ashcraft & Battaglia, 1978), by the linear trend between sum and RTs. It should be remembered that Ashcraft and Battaglia's proposal was mainly motivated by two findings. First, any counting hypothesis was rejected on the basis of the non-linear increase in RTs and the fact that 20 ms were judged too short an increment for counting, pointing towards a retrieval process. Second, this retrieval was assumed to take the form of a search on a table because RTs were related with the magnitude of the operands through the sum^2 predictor. However, we did not found an exponential but a linear increase in RTs.

If retrieval-based models including the hypothesis of the search through an additive mental table have difficulties in accounting for our results, a simpler and more plausible account becomes possible, which is that additions could be solved by a process of rapid scrolling through an easily accessible and overlearned representation stored in long-term memory. This representation could take the form of a mental number line or a verbal number sequence of the first numbers. Entering this line or sequence by its origin, the procedure would search for the position corresponding to the first operand, and then move forward by a number of steps corresponding to the second operand. Response times would reflect the distance from the origin to the value corresponding to the correct sum, accounting for the observed linear relationship between sum and RTs. This mechanism could be constructed by compiling basic activities rooted in counting and solving additions through a counting-all strategy, and could correspond to the compacted procedures or schemas evoked by Baroody (1994). Of course, we are aware of the fact that an increment of 20 ms per step has always been considered as ruling out any counting-based account of addition solving. Groen and Parkman (1972) themselves considered 20 ms as an unreasonable incrementing rate for adult's mental counting, arguing that the rate of silent counting in adults is about one number every 150 ms (Landauer, 1962). However, the compacted procedure would not necessarily rely on silent counting, but rather on the mental scanning of a portion of an ordered spatial or verbal representation. Indeed, empirical findings indicate that processing lists of items in this way can be surprisingly fast. For example, when studying the processes by which word lists are maintained in an active state for immediate serial recall in memory span tasks, Cowan, Saults, and Elliott (2002) estimated to about 40 ms the time needed to reactivate one item through a covert retrieval process. If ad hoc word lists constructed for the purpose of memory experiments and learned through a single exposure are accessed at a rate of 40 ms per item, it can be imagined that deeply rooted and automated lists such as the number line or the verbal number sequence can be accessed at a rate of 20 ms per item. Moreover, accessing up to three or four numbers in immediate succession, which was the maximum magnitude of the operands we used, would not exceed the capacities of the focus of attention as estimated by Cowan (2001). This renders possible the simultaneous representation in a single attentional focusing of a small portion of the number line (or the verbal number sequence) and the control of a small number of steps corresponding to the second operand when moving forward from the value corresponding to the first operand. The control of this small number of steps could rely on a process akin to subitizing, the limitations of which have already been accounted for by the limitation of the focus of attention (Cowan, 2001). In the same way as subitizing presents a small but reliable slope, this scrolling of the number line or the verbal numerical sequence could take longer as the number of steps increases, engendering the problem-size effect that we observed. Occurring in a single attentional focusing and relying on the mental scrolling of an overlearned sequence, this compacted procedure could be especially fast. If our hypothesis of a compacted procedure operating on the items held within a single focus of attention is correct, this means that adding operands larger than 4 would require a different type of procedure requiring several successive attentional focusing. This kind of procedure would probably be far slower than the compacted procedure, in the same way as, in enumerating collections, counting is slower and involves steeper slopes than subitizing (Mandler & Shebo, 1982). This latter hypothesis deserves empirical confirmation and further studies.

Our results suggest that, even with small operands, this compacted and highly automated procedure would be superseded by retrieval when solving tie problems for which the best predictor was not the sum but an interference-related index. However, we do not think that the faster responses observed on tie problems were due to the fact that retrievals are necessarily faster than algorithmic procedures, because tie problems benefit also from an encoding advantage for repeated operands that could underpin the observed difference. The preferential use of retrieval over the compacted procedure for tie problems could be due to the salience of the unique association between each small integer and the answer of its associated tie addition (i.e., 1–2, 2–4, 3–6) from which declarative knowledge such as "eight is two times four" or "two is half of four" could be derived.

Such a compacted procedure could correspond to the reconstructive strategy primed by the anticipated presentation of the additive sign in Fayol and Thevenot's (2012) study. It is worth to note that the hypothesis of a compacted procedure could also account for the individual differences observed. Barrouillet, Lépine, et al. (2008) suggested that the influence of working memory capacity on cognition is mediated by the impact of a basic general-purpose resource that affects each atomic step of cognition. The faster responses and flatter slopes observed with higher working memory capacities could result from a capacity to perform more quickly each step of the procedure. We also observed that individuals with the highest working memory capacity were no longer affected by the size of the first operand, whereas the size of the second operand had still an effect. This finding suggests some capacity to directly access the position corresponding to the first operand. This could result from a capacity to access simultaneously different locations or to a higher mastery in the manipulation of the number line. Finally, we observed that high and low working memory capacity individuals do not differ on their sensitivity to tie vs. non-tie problems. This suggests that the faster responses on tie problems do not depend on a cognitive process highly dependent on working memory capacity, corroborating the idea that these faster responses probably result from some peripheral process such as encoding rather than from deeper processes involving central activities.

A last question is in what extent such a compacted and automated procedure would differ from a search through a table, as suggested by Ashcraft and Battaglia (1978). It could be argued that the two hypotheses are akin because the search through a table would involve the same scrolling of a "number line-like" representation to access the target intersection between line and column. However, a main difference is that the procedure we describe does not require the storage and retrieval of problem-answer associations, but only uses the number line or the verbal numerical sequence and their numerical properties that are discovered early in development (Fuson, Richards, & Briars, 1982). This is not to say that such associations do not exist and that they are never used. As we have seen, there were participants who exhibited flat RT patterns corresponding to what a retrieval hypothesis would predict, and they were faster than the others. In the same way, retrieval remains the most plausible hypothesis for tie problem solving. However, it is also possible that the compacted procedure is sufficiently fast to efficiently compete with retrieval and to be used by a large majority of individuals to solve even the smallest additive problems.

To conclude, precise measures on a large sample of participants solving a restricted range of very small additive problems revealed a pattern of RTs that does not fit very well with current accounts of the problem-size effect. Indeed, increase in RTs seems strongly related to the magnitude of the operands. Considering the restricted range of problems we studied, our hypothesis of a compacted and automated procedure should be seen as a speculative attempt. Nonetheless, our results echo and reinforce previous proposals by Baroody (1994) and meshes with recent findings from Fayol and Thevenot (2012) that are difficult to reconcile with the received view that small additions are solved by direct retrieval.

References

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology, 6*, 451–474.

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., & 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.

Baroody, A. J. (1994). An evaluation of evidence supporting fact-retrieval models. *Learning and Individual Differences, 6*, 1–36.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 133*, 83–100.

Barrouillet, P., & Lépine, R. (2005). Working memory and children's use of retrieval to solve addition problems. *Journal of Experimental Child Psychology, 91*, 183–204.

Barrouillet, P., Lépine, R., & Camos, V. (2008). 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., Mignon, M., & Thevenot, C. (2008). Strategies in subtraction problem solving in children. *Journal of Experimental Child Psychology, 99*, 233–251.

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. (1999). An embedded processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge: Cambridge University Press.

Cowan, N., Saults, J. S., & Elliott, E. M. (2002). The search for what is fundamental in the development of working memory. *Advances in Child Development and Behavior, 29*, 1–49.

Fayol, M., & Thevenot, C. (2012). The use of procedural knowledge in simple addition and subtraction problems. *Cognition, 123*, 392–403.

Ferrand, L., Brysbaert, M., Keuleers, E., New, B., Bonin, P., Meot, A., et al. (2011). Comparing word processing times in naming, lexical decision, and progressive demasking: Evidence from chronolex. *Frontiers in Psychology, 2*, 306.

Fuson, K. C., Richards, J. J., & Briars, D. J. (1982). The acquisition and elaboration of the number word sequence. In J. C. Brainerd (Ed.). *Progress in cognitive development. Children's logical and mathematical cognition* (Vol. 1, pp. 33–92). New York: Springer-Verlag.

Geary, D. C., Bow-Thomas, C., Liu, F., & Siegler, R. S. (1996). Development of arithmetical competencies in Chinese and American children. *Child Development, 67*, 2022–2044.

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.

Landauer, T. K. (1962). Rate of implicit speech. *Perceptual and Motor Skills, 15*, 646.

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

Lovett, M. C., Reder, L. M., & Lebière, C. (1999). Modeling working memory in a unified architecture: An ACT-R perspective. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 135–182). Cambridge: Cambridge University Press.

Mandler, G., & Shebo, B. J. (1982). Subitizing: an analysis of its component processes. *Journal of Experimental Psychology, 111*, 1–22.

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.

Siegler, R. S., & Shrager, J. (1984). Strategic choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229–293). Hillsdale: Erlbaum.

Thevenot, C., Barrouillet, P., & Fayol, M. (2001). Algorithmic solution of arithmetic problems and operand-answer associations in long term memory. *Quarterly Journal of Experimental Psychology, 54*(2), 599–611.

Woltz, D. J. (1988). An investigation of the role of working memory in procedural skill acquisition. *Journal of Experimental Psychology: General, 117*, 319–331.

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.