Journal of Experimental Child Psychology 234 (2023) 105710

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

The development of simple addition problem solving in children: Reliance on automatized counting or memory retrieval depends on both expertise and problem size

Céline Poletti [a], Andrea Díaz-Barriga Yáñez [b], Jérôme Prado [b,*,1], Catherine Thevenot [a,*,1]

[a] Institut de Psychologie, Université de Lausanne, CH-1015 Lausanne, Switzerland
[b] Lyon Neuroscience Research Center (CRNL), INSERM U1028–CNRS UMR5292, University of Lyon, 69675 Bron Cedex, France

A R T I C L E   I N F O

Article history:
Received 29 January 2023
Revised 5 April 2023
Available online 6 June 2023

Keywords:
Numerical cognition
Arithmetic
Addition
Memory network
Horse race model

A B S T R A C T

In an experiment, 98 children aged 8 to 9, 10 to 12, and 13 to 15 years solved addition problems with a sum up to 10. In another experiment, the same children solved the same calculations within a sign priming paradigm where half the additions were displayed with the "+" sign 150 ms before the addends. Therefore, size effects and priming effects could be considered conjointly within the same populations. Our analyses revealed that small problems, constructed with addends from 1 to 4, presented a linear increase of solution times as a function of problem sums (i.e., size effect) in all age groups. However, an operator priming effect (i.e., facilitation of the solving process with the anticipated presentation of the "+" sign) was observed only in the group of oldest children. These results support the idea that children use a counting procedure that becomes automatized (as revealed by the priming effect) around 13 years of age. For larger problems and whatever the age group, no size or priming effects were observed, suggesting that the answers to these problems were already retrieved from memory at 8 to 9 years of age. For this specific category of large

* Corresponding authors.
E-mail addresses: jerome.prado@univ-lyon1.fr (J. Prado), catherine.thevenot@unil.ch (C. Thevenot).
[1] These authors share senior authorship.

https://doi.org/10.1016/j.jecp.2023.105710
0022-0965/

problems, negative slopes in solution times demonstrate that retrieval starts from the largest problems during development. These results are discussed in light of a horse race model in which procedures can win over retrieval.

Introduction

At the beginning of arithmetic learning, it is undisputed that young children solve single-digit addition problems using counting strategies (e.g., Carpenter & Moser, 1984; Siegler & Shrager, 1984). For example, they can count in ones starting from the largest addend when solving a problem such as 4 + 7 (i.e., 7 + 1 = 8 + 1 = 9 + 1 = 10 + 1 = 11) (e.g., Groen & Parkman, 1972). They might also solve problems by decompositions and derived-fact strategies (e.g., 4 + 7 is 7 + 3 = 10 + 1) (e.g., Dowker, 2015). These initial strategies are easy to observe or infer from overt behaviors such as finger or verbal counting and slow solution times (e.g., Siegler & Robinson, 1982). However, with development and practice, strategies become increasingly internalized and fast, which also makes them increasingly difficult to identify (Campbell & Thompson, 2012; Thevenot et al., 2015). This partly explains why researchers have not reached a consensus yet concerning the way initial counting strategies evolve until expertise.

According to retrieval models, each time a problem is solved by reconstructive strategies such as counting and decomposition, an association between operands and answers is created. After repetitive practice or, in other words, once the association is strong enough, the problem can be solved by retrieval from long-term memory (e.g., Ashcraft, 1992; Logan, 1988a, 1988b). Retrieval is largely thought to be the most efficient and fastest solving strategy (e.g., Groen & Parkman, 1972; Kaye et al., 1986; Siegler & Robinson, 1982). Therefore, it is supposed to be increasingly used by individuals (e.g., Siegler & Shrager, 1984) until it becomes the dominant strategy starting around Grade 4 (Ashcraft & Battaglia, 1978) for single-digit addition problems. According to these models, expertise is reached when individuals are able to directly retrieve the problem answer from memory without relying on counting (e.g., Ashcraft, 1992; Chen & Campbell, 2019; Logan, 1988a, 1988b; Siegler & Shrager, 1984).

In opposition to retrieval models, proponents of counting models defend the idea that there is not a necessary shift from counting to retrieval during arithmetic learning. Instead, the development of arithmetic expertise might consist in an acceleration of counting procedures until their automatic execution (e.g., Baroody, 1983, 1984; Svenson, 1985; Thevenot et al., 2016; Uittenhove et al., 2016). Therefore, in contrast to retrieval models (Ashcraft, 1992; Campbell, 1995; Siegler & Jenkins, 1989), counting models consider that the development of arithmetic skills is more quantitative (i.e., it involves an acceleration of procedures) than qualitative (i.e., it relies on the successive use of different, sometimes overlapping, strategies) (Siegler, 1996). More precisely, the automatized counting procedure theory suggests that slow counting algorithms that are implemented by young children with the aid of objects or fingers are progressively internalized, resulting in mental procedures (e.g., Baroody, 1983). The main idea is that these algorithms, which are initially slow and require awareness and high cognitive control, could be reinforced each time they are completed until they become so fast that they no longer need verbalization of each counting step. At this automatization stage, only the end product of the algorithm and not each of its steps is under conscious control of the problem solvers (Logan, 2018; Uittenhove et al., 2016).

Although the counting theory is still debated (see, e.g., Chen & Campbell, 2018, and Thevenot & Barrouillet, 2020), it has received recent support from behavioral, neuropsychological, and brain imaging studies (e.g., Evans & Ullman, 2016; Mathieu et al., 2018b; McDougle et al., 2022; Pinheiro-Chagas et al., 2017). Evidence of automatized counting procedures for addition has been first provided by the sign (or operator) priming paradigm, in which the sign of the operation to be solved is presented shortly before the operands (i.e., classically 150 ms before). The seminal studies using this paradigm in adults showed that priming the "+" sign accelerates addition problem solving, whereas the same manipulation with the "×" sign has no effect on multiplication (Fayol & Thevenot, 2012; Roussel et al., 2002). These results have also been replicated in older adults (Thevenot et al., 2020b). They suggest that a procedure, independent of the operands, may be activated as soon as the "+" sign is presented. This does not appear to be the case when the "×" sign is presented. Therefore, it was concluded that addition problems are mainly solved by counting procedures, whereas multiplication problems (because they are learned by rote at school) are mainly solved by retrieval of the results from long-term memory.

The nature of this procedure has been specified subsequently in a series of experiments showing that addition problem solving is facilitated when the second operand of the problem is presented on the right side on a computer screen rather than on the left side (Díaz-Barriga Yáñez et al., 2020; Mathieu et al., 2016). Therefore, solving addition problems is facilitated when the attention of participants is drawn to their right attentional field. This suggests that addition problems may be solved through displacements on a number line from left to right and that the presentation of the second operand in the direction of the displacement eases out the solving process. These findings allowed an interpretation of the nature of the procedures that is primed by the "+" sign for additions: They might correspond to the preactivation of the mental number line on which the displacements are made (as well as the mental preparation for such displacements).

However, developmental studies indicate that automatized counting procedures might emerge relatively late. For instance, addition sign priming effects are typically not observable before 13 years of age (Mathieu et al., 2018a; Poletti et al., 2021). This is consistent with the fact that spatial facilitation effects associated with perceiving the second operand in the right visual field also emerge at age 13 for addition problems (Díaz-Barriga Yáñez et al., 2020). This suggests that children might not be able to convoke automatized counting procedures to solve addition problems before age 13.

A central point of the automatized counting theory is that automatized procedures could also be limited to very small addition problems involving operands from 1 to 4, that is, within the subitizing range (Barrouillet & Thevenot, 2013; Uittenhove et al., 2016). It is possible that, beyond this operand range, the answers to problems with a sum up to 10 are retrieved from long-term memory. This iconoclast assumption that larger addition problems could be more frequently solved by retrieval than smaller addition problems stems from close examination of the problem sum effect on solution times in adults and children (Bagnoud et al., 2021a; Uittenhove et al., 2016). Indeed, whereas a linear increase in solution times is observed for problems with a sum from 3 to 7, there is no longer an increase in solution times for problems with a sum of 8, 9, or 10 (see Fig. 2 in Bagnoud et al., 2021a, or Fig. 1 in Uittenhove et al., 2016). The lack of association between problem sums and solution times is classically interpreted as evidence of the use of retrieval strategies (e.g., Dewi et al., 2021a; Logan, 1988a, 1988b; Logan & Klapp, 1991; Thevenot et al., 2020a). If this interpretation is admitted, then an increase in solution times reflects the use of either counting or a mix between counting and retrieval (e.g., Compton & Logan, 1991).

Therefore, we are left in a situation where individuals from 13 years of age could solve very small addition problems by automatized procedures, whereas larger problems could be solved using retrieval. This hypothesis was tested in the current study by extending Poletti et al.'s (2021) research and studying the sign priming effect with larger problems. Indeed, whereas Poletti et al. used only small additions with operands up to 4, children in the current study were presented with all non-tie problems with a sum up to 10. This would allow us not only to replicate Poletti et al.'s results that small addition problems are primed by the "+" sign from age 13 but also to examine whether larger addition problems are subjected to a sign priming effect. If we are right in assuming that, contrary to very small problems, the answers to larger problems with a sum up to 10 are retrieved from memory, no priming effect should be observed for this category of problems irrespective of the age of children involved in the experiment (i.e., children aged 8–9, 10–12, and 13–15 years).

In addition to sign priming effects, the distribution of solution times depending on the size of the problem sums was also examined. Here we expected to replicate Bagnoud et al.'s (2021a) results showing that whereas an increase in solution times is associated with very small problems, a flat distribution of solution times characterizes problems with sums of 8, 9, and 10. Moreover, examination of the size and sign priming effects conjointly within the same populations of children would be possible for the first time. This conjoint examination should also allow for the investigation of potential correlations between these effects. Indeed, it is so far not clear whether the size of the priming effect, not only its plain presence or absence, is related to individuals' expertise. In addition to size effects, we examined whether a more direct measure of arithmetic fluency (an index of children's expertise in arithmetic) correlates with size and priming effects. Furthermore, whereas the presence of a priming effect is commonly interpreted as the preactivation of an automatized procedure, the lack of a priming effect can be interpreted as either the use of retrieval or the use of conscious solving procedures. Conjoint examination of size and priming effects should allow us to disambiguate these interpretations. Indeed, a lack of priming effect associated with large size effects would be confidently interpreted as the use of nonautomatized conscious arithmetic procedures because these are time-consuming and necessarily associated with significant size effects. In contrast, a lack of priming associated with negligible, or no size effect would be interpreted as the use of retrieval.

To sum up and complete our predictions, according to both the automatized counting and retrieval theories, conscious counting procedures would be used by younger children. Therefore, priming effects should not be observed in younger children. However, the theories differ with regard to predictions with older children. The automatized counting model predicts that the priming effect of the "+" sign for small problems should be significant in 13- to 15-year-olds because these children have automatized addition procedures. According to the retrieval theory, retrieval would be systematically used by older children to solve addition problems. Therefore, priming effects should not be observed in older children, much like in younger children. Concerning size effects, a larger increase in solution times for small problems than for larger problems would be incompatible with the retrieval theory, particularly in 13- to 15-year-olds. This is because these children are supposed to massively retrieve the results of simple addition problems with sums up to 10 from memory. As already explained, size effects in retrieval models are mainly explained by frequency and interference effects, which increase monotonically with the size of the problems (Ashcraft & Christy, 1995; Zbrodoff, 1995).

Method

Given that the participants and stimuli were the same in Experiment 1 (priming effects) and Experiment 2 (size effects), they are described first. However, the procedures and results, which are specific to each experiment, are presented in sections devoted to each of them.

Participants

A total of 105 French-speaking Swiss children took part in this study. The sample was composed of 39 children aged 8 to 9 years (M = 8.94 years, SD = 0.30; 25 girls), 42 children aged 10 to 12 years (M = 11.48 years, SD = 0.82; 19 girls), and 24 children aged 13 to 15 years (M = 14.38 years, SD = 0.76; 10 girls). None of the children suffered from learning disabilities.

Our study was conducted following the principles of the Declaration of Helsinki. Parental written consent was collected for each child. More precisely, parents gave their consent for their children's participation in our study and for the inclusion of their results in our analyses. They were informed that their children's results would be fully anonymized.

Stimuli

Children were instructed to solve arithmetic problems orally as quickly and accurately as possible. Children solved all the problems with a sum up to 10 with operands from 1 to 9 (i.e., 45 problems), and each problem was presented three times (i.e., total of 135 problems). However, tie problems and problems involving 1 were not considered in our analyses. Indeed, it is largely accepted that tie problems are solved by retrieval by the end of first grade (Bagnoud et al., 2021a; LeFevre et al., 1996). Problems involving 1 are also suspected to be solved by a rule and, therefore, are not necessarily sensitive to priming effects (Bagnoud et al., 2021b; Baroody, 1985; Grabner et al., 2022). Moreover, these problems can be associated with very hectic solution times that can pollute size effects (Bagnoud et al., 2021a). The remaining problems were split into two categories. Small problems corresponded to problems involving operands from 2 to 4, whereas larger problems (hereafter *large problems*) contained at least one operand larger than 4 (see Table 1 for the list of addition problems considered in our analyses). One characteristic of this classification is that two problems with the same sum may belong to two different categories. Indeed, whereas 3 + 4 belongs to the small category, 5 + 2 belongs to the large category. Therefore, and contrary to previous classifications (e.g., Fayol & Thevenot, 2012), the problems were not differentiated *stricto sensu* based on their sum but instead (as noted above) on the fact that the operands both could be subitized or not.

Arithmetic fluency test

Children's arithmetic skills were assessed using a subtest of Woodcock–Johnson III. In this paper-and-pencil test, children are asked to solve additions, subtractions, and multiplications presented in columns involving operands from 0 to 10. They have 3 min to solve as many problems as possible out of the 160 problems presented. The final score is the number of correct answers provided.

Experiment 1: Sign priming paradigm

Procedure

The pairs of digits were presented in the addition and multiplication conditions. It was necessary to include multiplications in our design because priming effects cannot be observed when only one operation is presented. Indeed, in this case the potential procedure used to solve the additions would be activated all along the task and could not be preactivated. Still, only addition sign priming effects are presented here because multiplication was not the object of our research.

The arithmetic sign was presented either 150 ms before the operands (i.e., −150-ms stimulus onset asynchrony [SOA] condition) or at the same time as the operands (i.e., null SOA condition). We constructed 540 problems in total (i.e., 135 pairs of digits × 2 operations × 2 SOA). Because it would have been difficult for all children to solve such a large set of problems, the material was divided into four sets of 135 problems and each participant was tested on only one of these four sets. Therefore, the entirety of the material was presented after 4 children were tested. Problems were randomly presented within each set.

The experiment was run under DMDX software (Forster & Forster, 2003). Vocal responses were recorded with a voice key and individually checked offline for accuracy using CheckVocal software (Protopapas, 2007). CheckVocal was also used to manually adjust the latencies recorded by DMDX if necessary. More precisely, for each response recorded, CheckVocal allows for the visualization of the sound played out through a waveform. When, despite precalibration of the voice key for sensitivity, the onset of the response given by participants is not accurately detected, the timing mark can be manually placed on the onset of the sound waveform. This checking and possible manual readjustments ensure a measure of solution time with 1-ms precision.

Each trial began with the presentation of a 500-ms fixation signal "|", followed by the presentation of the problem (i.e., the arithmetic sign followed by the operands in the 150-ms SOA condition or the arithmetic sign and the operands simultaneously in the null SOA condition; Fig. 1). The problem was displayed on the screen until a verbal response onset was detected by the voice key. Solution times

Table 1
List of additions considered in our analyses.

Small operands	Large operands		
2 + 3	2 + 5	3 + 7	6 + 2
2 + 4	2 + 6	4 + 5	6 + 3
3 + 2	2 + 7	4 + 6	6 + 4
3 + 4	2 + 8	5 + 2	7 + 2
4 + 2	3 + 5	5 + 3	7 + 3
4 + 3	3 + 6	5 + 4	8 + 2

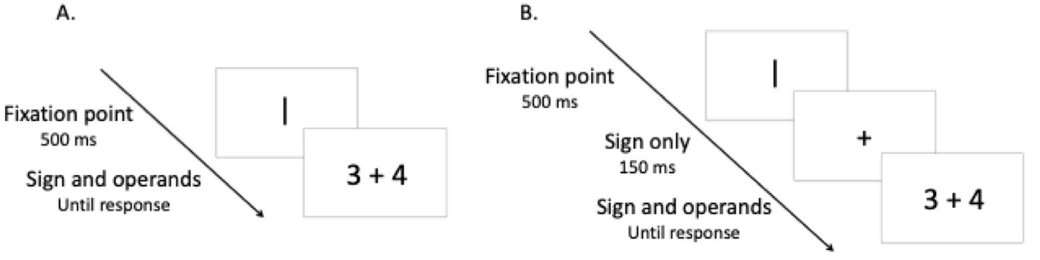

Fig. 1. Examples of trial sequences for addition problem in the null stimulus onset asynchrony (SOA) condition (A) and the –150-ms SOA condition (B).

corresponded to the time elapsed between the presentation of the problem in its whole and voice key activation. To familiarize children with the task and allow the experimenter to test the voice key sensitivity, 6 warm-up problems (i.e., 3 additions and 3 multiplications) were presented before the experimental phase. To avoid excessive fatigue, four breaks were proposed during the course of the experiment. Each child was tested individually in a quiet room within the school, and the completion of the task took about 20 min.

Results

The data sets that were generated and analyzed in Experiment 1 are available in the Open Science Framework (OSF) repository https://osf.io/fpnav/?view_only=1bfa02854f0a4210871add96630dc1d7. Seven children (3 children aged 8–9 years, 3 children aged 10–12 years, and 1 child aged 13–15 years) were excluded from the analyses because their mean solution times differed by more than 2 standard deviations from the mean solution times of their age group. Therefore, the analyses were conducted on a total of 98 participants.

Because our critical predictions centered around the presence (or absence) of a priming effect in older (but not younger) children, we present the results of 2 (Problem Size: small or large) × 2 (SOA: null or negative) analyses of variance (ANOVAs) conducted separately for each age group.

Percentages of errors

The analysis on percentages of errors was carried out on 91% of the whole set of data because, due to technical problems, no response was recorded for 9% of trials. Overall, children performed well on the task, with less than 5% of errors (4.3% exactly; see Table 2). A 2 (Problem Size: small or large) × 2 (SOA: null or negative) ANOVA with both factors as repeated measures was performed on percentages of errors for each age group.

Children aged 8 to 9 years. The main effects of problem size ($F < 1$, $\eta^2_p = .00$, $p = .79$) and SOA ($F < 1$, $\eta^2_p = .00$, $p = .74$) and the interaction between these two factors ($F < 1$, $\eta^2_p = .01$, $p = .53$) were not significant. Post hoc analyses confirmed that there was no significant priming effect of the arithmetic sign on percentages of errors, whether the problem was small ($F < 1$, $\eta^2_p = .00$, $p = .82$) or large, $F(1, 35) = 1.08$, $\eta^2_p = .03$, $p = .31$.

Children aged 10 to 12 years. The main effects of problem size ($F < 1$, $\eta^2_p = .00$, $p = .70$) and SOA ($F < 1$, $\eta^2_p = .00$, $p = .87$) were not significant. The Problem Size × SOA interaction did not reach significance either, $F(1, 38) = 2.22$, $\eta^2_p = .06$, $p = .14$. Post hoc analyses confirmed that there was no significant priming effect of the arithmetic sign on percentages of errors, whether the problem was small ($F < 1$, $\eta^2_p = .02$, $p = .39$) or large, $F(1, 38) = 1.93$, $\eta^2_p = .05$, $p = .17$.

Children aged 13 to 15 years. The main effects of problem size, $F(1, 22) = 1.84$, $\eta^2_p = .08$, $p = .19$, and SOA, $F(1, 22) = 1.49$, $\eta^2_p = .06$, $p = .23$, were not significant, but the Problem Size × SOA interaction reached

Table 2
Mean solution times (in milliseconds) and percentages of errors as a function of age group, problem size, and SOA.

	8- to 9-year-olds		10- to 12-year-olds		13- to 15-year-olds	
Condition	Small	Large	Small	Large	Small	Large
Percentages of errors						
Null SOA	4.4 (10.4)	3.1 (5.8)	3.9 (9.5)	5.2 (6.8)	8.8 (14.6)	2.7 (5.1)
Negative SOA	3.9 (10.7)	4.5 (7.6)	5.5 (10.3)	3.2 (5.3)	2.8 (7.5)	4.1 (5.1)
Solution times (ms)						
Null SOA	1978 (571)	2036 (525)	1574 (491)	1708 (505)	1277 (299)	1290 (271)
Negative SOA	1914 (680)	2069 (619)	1539 (635)	1667 (561)	1203 (269)	1262 (289)
Priming effects	+64 (83)	−34 (59)	+35 (52)	+41 (28)	+74* (33)	+28 (16)

Note. Standard deviations are in parentheses. Priming effects correspond to the difference between solution times in the null and negative stimulus onset asynchrony (SOA) conditions.
* $p < .05$.

significance, $F(1, 22) = 4.29$, $\eta_p^2 = .16$, $p = .05$, showing that the priming effect (i.e., the positive difference between the null and negative SOA conditions) was larger for small problems (+6.0%) than for large problems (−1.4%). However, post hoc analyses showed that priming effects were not significant, whether the problem was small, $F(1, 22) = 2.99$, $\eta_p^2 = .12$, $p = .10$, or large, $F(1, 22) = 1.17$, $\eta_p^2 = .05$, $p = .29$.

Solution times

A 2 (Problem Size: small or large) × 2 (SOA: null or negative) ANOVA with both factors as repeated measures was performed on solution times for each age group (see Table 2).

Children aged 8 to 9 years. A significant main effect of problem size revealed that children tended to be faster to solve small addition problems than large ones (−107 ms), $F(1, 35) = 3.08$, $\eta_p^2 = .08$, $p = .09$. However, the main effect of SOA ($F < 1$, $\eta_p^2 = .00$, $p = .75$) and the interaction between problem size and SOA ($F < 1$, $\eta_p^2 = .02$, $p = .38$) were not significant. Post hoc analyses confirmed that there was no significant priming effect of the arithmetic sign, whether the problem was small ($F < 1$, $\eta_p^2 = .02$, $p = .45$) or large ($F < 1$, $\eta_p^2 = .01$, $p = .57$).

Children aged 10 to 12 years. A significant main effect of problem size revealed that children were faster to solve small addition problems than large ones (−130 ms), $F(1, 38) = 5.31$, $\eta_p^2 = .12$, $p = .03$. However, the effect of SOA, $F(1, 38) = 1.32$, $\eta_p^2 = .03$, $p = .26$, and the interaction between problem size and SOA ($F < 1$, $\eta_p^2 = .00$, $p = .91$) were not significant. Post hoc analyses confirmed that there was no significant priming effect of the arithmetic sign, whether the problem was small ($F < 1$, $\eta_p^2 = .01$, $p = .51$) or large, $F(1, 38)$, $\eta_p^2 = .05$, $p = .15$.

Children aged 13 to 15 years. There was no significant effect of problem size, $F(1, 22) = 1.04$, $\eta_p^2 = .05$, $p = .32$. However, there was an effect of SOA, showing that children were faster in the negative condition than in the null SOA condition (+51 ms), $F(1, 22) = 6.47$, $\eta_p^2 = .23$, $p = .02$. The Problem Size × SOA interaction was not significant, $F(1, 22) = 2.02$, $\eta_p^2 = .08$, $p = .17$. However, post hoc analyses revealed that the effect of SOA was significant for small problems (+74 ms), $F(1, 22) = 5.11$, $\eta_p^2 = .19$, $p = .03$, but not for larger ones (+28 ms), $F(1, 22) = 2.88$, $\eta_p^2 = .12$, $p = .11$.

Discussion

The results of this experiment show facilitation effects of the anticipated presentation of the "+" sign for small problems in the group of older children (i.e., 13- to 15-year-olds) on both percentages of errors and solution times. Note, however, that priming effects on the percentages of errors in older children cannot be considered as supporting our hypotheses. Indeed, this effect is not due to a reduction of errors in the negative SOA condition but instead to an increase of errors in the null SOA condition for small problems. A look at the individual data revealed that this result was due to the fact that, for small problems, children sometimes mistakenly gave the result of the multiplication instead of the addition of the operands. Of course, when the "+" sign was presented before the operand, this impulsive behavior was limited.

Therefore, as in our previous studies, our discussion is based only on response times. As explained in our Introduction, arithmetic sign priming effects can be explained by the use of automatized procedures. Thus, our results suggest, in line with previous observations (Poletti et al., 2021) and in accordance with prior conclusions using different paradigms (Díaz-Barriga Yáñez et al., 2020; Mathieu et al., 2018) that the use of automatized counting procedures for small addition problems is the dominant strategy from 13 years of age.

Our next steps in the current article are to examine size effects in the same populations of children and, finally, to interpret priming and size effects, one in light of the other. As already explained in our Introduction, this approach allows the interpretation of a lack of priming effect, which can reflect either the use of retrieval or the use of conscious, diverse, and costly counting procedures that cannot be primed in 150 ms.

Experiment 2: Size effects

Procedure

Children were asked to solve simple additions and to give their responses orally as quickly and accurately as possible. The same 135 possible combinations of one-digit numbers as in Experiment 1 were used to construct addition problems (see Table 1). The task was also designed using DMDX software. Vocal responses were recorded with a voice key and individually checked offline for accuracy using CheckVocal software.

Each trial began with a fixation signal "|" displayed at the center of a computer screen for 500 ms. Then, the addition problem was presented in its entirety at the center of the screen and remained on the screen until a response was given orally by the children. Children were presented with three blocks of problems, each containing the 45 problems described in the Method section. The problems were presented randomly in each block. To familiarize children with the task and allow the experimenter to test the voice key sensitivity, 4 warm-up addition problems were presented before the experimental phase. To avoid excessive fatigue, one break was proposed in each block. As in Experiment 1, each child was tested individually in a quiet room within the school, and the task took about 15 min to complete.

Results

The data sets that were generated and analyzed in Experiment 2 are available in the OSF repository (https://osf.io/fpnav/?view_only=1bfa02854f0a4210871add96630dc1d7). To examine size and priming effects conjointly, the data sets of the 7 children who were excluded from the analyses in Experiment 1 were also excluded from the current analyses.

For this experiment, and contrary to Experiment 1, specific effects depending on age groups were not expected. Therefore, we classically ran ANOVAs including this variable.

Percentages of errors

The analysis on percentages of errors was carried out on 95% of the whole set of data because no response was recorded for 5% of trials due to technical errors. Overall, children performed very well on the task, making about 4.1% of errors (Table 3). An ANOVA on this variable was conducted with age group (8- to 9-year-olds, 10- to 12-year-olds, or 13- to 15-year-olds) as a between factor and problem size (small or large) as a within factor. The main effects of age group, $F(2, 95) = 2.33$, $\eta_p^2 = .05$, $p = .10$, and problem size ($F < 1$, $\eta_p^2 = .01$, $p = .38$) and the interaction between these two factors ($F < 1$, $\eta_p^2 = .01$, $p = .81$) were not significant.

Table 3
Mean solution times (in milliseconds) and percentages of errors as a function of age group and problem size *Note.* Standard deviations are in parentheses.

Problem size	8- to 9-year-olds	10- to 12-year-olds	13- to 15-year-olds
Percentages of errors			
Small problems	3.8 (7.9)	3.3 (4.9)	5.9 (7.1)
Large problems	2.8 (3.4)	3.3 (3.6)	5.3 (5.3)
Solution times (ms)			
Small problems	1720 (516)	1316 (446)	1159 (329)
Large problems	1768 (454)	1423 (433)	1167 (246)

Solution times

This analysis was carried out on correctly solved problems only (i.e., 96% of the trials analyzed in the previous section). In addition to the 5% of trials already discarded because of technical errors, 4% of outliers corresponding to responses below 200 ms and more than 2 standard deviations away from the participants' mean were also discarded from the analysis on solution times.

To analyze the difference between small and large problems, we conducted an ANOVA on solution times with age group (8- to 9-year-olds, 10- to 12-year-olds, or 13- to 15-year-olds) as a between factor and problem size (small or large) as a within factor (Table 3). There was a significant main effect of problem size, $F(1, 95) = 5.96$, $\eta_p^2 = .06$, $p = .02$, with shorter solution times for small problems (1398 ms) than for large problems (1453 ms). The results also revealed a main effect of age group, $F(2, 95) = 15.28$, $\eta_p^2 = .24$, $p < .001$. Post hoc analyses indicated that solution times in 8- to 9-year-olds (1744 ms) were significantly longer than those in 10- to 12-year-olds (1369 ms), $t(95) = 3.91$, $p < .001$, and in 13- to 15-year-olds (1163 ms), $t(95) = 5.25$, $p < .001$. The difference in solution times between 10- to 12-year-olds and 13- to 15-year-olds (−206 ms) was marginally significant, $t(95) = 1.89$, $p = .06$. The Age Group × Problem Size interaction was not significant, $F(2, 95) = 1.63$, $\eta_p^2 = .03$, $p = .20$.

Size effects

Small problems. Solution times for small problems followed a linear pattern. Therefore, it was possible to fit them with a linear regression for each participant, with the problem sum as the predictor. The slopes were extracted and Student's *t* tests were performed to assess whether they were statistically different from 0 in each age group. This was indeed the case, with $t(35) = 6.95$, $p < .001$, $d = 1.16$, and a slope of 282 ms for 8- to 9-year-olds, $t(38) = 4.01$, $p < .001$, $d = 0.64$, and a slope of 155 ms for 10- to 12-year-olds, and $t(22) = 2.45$, $p = .02$, $d = 0.51$, and a slope of 98 ms for 13- to 15-year-olds (see Fig. 2).

Large problems. In sharp contrast to small problems and as can be seen in Fig. 2, at least from sum to 8, large problems did not exhibit the classical problem size effect. To clarify this point, we performed an ANOVA with age group as a between factor and problem sum (sum to 7, sum to 8, sum to 9, and sum to 10 problems) as a within factor. The analysis indicated that sum to 10 problems were solved faster than sum to 7 [−153 ms, $t(95) = 4.49$, $p < .001$], sum to 8 [−223 ms, $t(95) = 6.80$, $p < .001$], and sum to 9 problems [−179 ms, $t(95) = 6.63$, $p < .001$]. Moreover, sum to 9 problems were solved tendentially faster than sum to 8 problems [−44 ms, $t(95) = 1.67$, $p = .09$]. This difference was especially due to the group of 13- to 15-year-olds for whom the difference approached significance [−104 ms, $t(95) = 1.96 p = .05$]. Concerning the difference in solution times between sum to 7 and sum to 8 problems, it was significant in 8- to 9-year-olds [−114 ms, $t(95) = 2.52 p = .01$] but not in 10- to 12-year-olds [−27 ms, $t(95) = 0.62$, $p = .54$] or in 13- to 15-year-olds [−70 ms, $t(95) = 1.23$, $p = .22$].

Furthermore, it appeared that, at least for the younger group of children, problems with a sum to 7 were solved quicker when they belonged to the large category of problems rather than the small category. This was confirmed by a Student's *t* test, $t(35) = 2.77$, $p = .01$, $d = 0.46$. For 10- to 12-year-olds and 13- to 15-year-olds, solution times for large and small sum to 7 problems were not statistically different ($p = .93$ and $p = .51$, respectively).

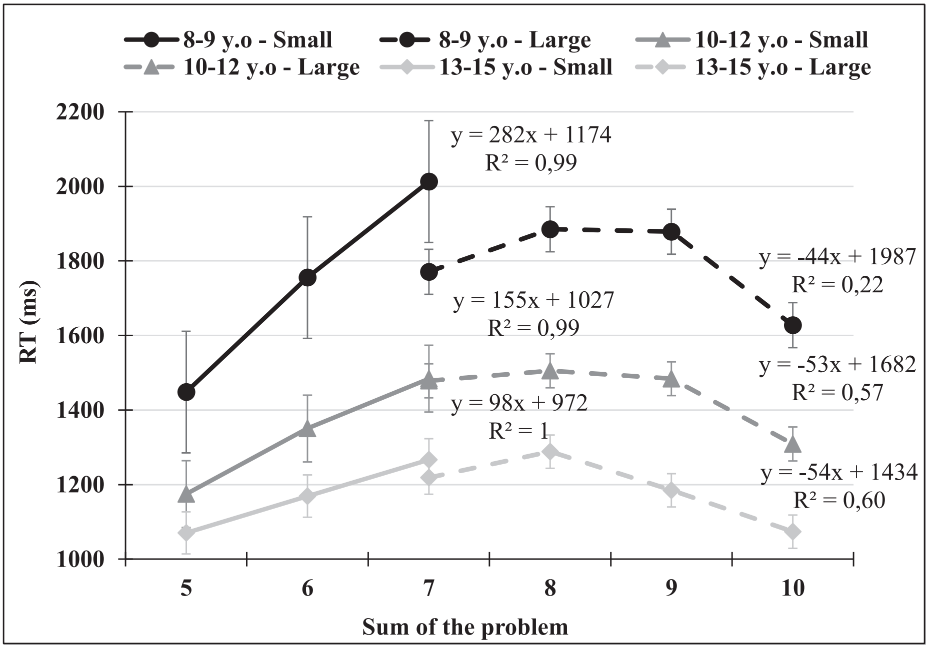

Fig. 2. Mean solution times (with standard errors) for each age group and each problem size according to the sum of the problem. y.o, years old.

Comparisons between small and larger problems. Student's *t* tests were performed to assess whether slopes associated with small problems were statistically higher than those associated with large problems. This was indeed the case when all age groups were considered together, $t(97) = 8.36$, $p < .001$, $d = 0.84$, and also when each age group was considered separately, $t(35) = 6.91$, $p < .001$, $d = 1.15$ for 8- to 9-year-olds, $t(38) = 4.65$, $p < .001$, $d = 0.75$ for 10- to 12-year-olds, and $t(22) = 2.82$, $p = .01$, $d = 0.59$ for 13- to 15-year-olds.

Experiments 1 and 2

Correlational analyses

Correlation analyses between arithmetic fluency scores, size, and priming effects for small problems and large problems (Table 4) were performed for the full sample of children and for each age group.

Concerning small problems and the full sample of children, there was a negative correlation between arithmetic fluency scores and the size effect, revealing that the size effect increases as children's arithmetic fluency decreases ($r = -.448$, $p < .001$). When the age of children was entered as a covariate in the analysis, the correlation remained significant ($r = -.344$, $p < .001$). This correlation was observed in the two groups of younger children ($r = -.477$, $p = .01$ and $r = -.402$, $p = .01$ for 8- to 9-year-olds and 10- to 12-year-olds, respectively). It was still significant when controlling for the age of children ($r = -.338$, $p = .05$ and $r = -.465$, $p = .01$ for 8- to 9-year-olds and 10- to 12-year-olds, respectively). Moreover, in 10- to 12-year-olds, there was a negative correlation between size effects and priming effects, revealing that larger size effects were related to smaller priming effects ($r = -.527$, $p < .001$).

Table 4
Correlations between arithmetic fluency scores, size effect in Experiment 2, and priming effect in Experiment 1 for small and large problems in the full sample of children (N = 98), 8- to 9-year-olds (n = 36), 10- to 12-year-olds (n = 39), and 13- to 15-year-olds (n = 23)

	Full sample					8- to 9-year-olds					10- to 12-year-olds					13- to 15-year-olds				
Variable	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1. Small problem size effects	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
2. Small problem priming effects	−.116	–	–	–	–	.169	–	–	–	–	−.527***	–	–	–	–	−.248	–	–	–	–
3. Large problem size effects	−.212*	.000	–	–	–	−.218	.161	–	–	–	−.187	−.221	–	–	–	−.382	−.005	–	–	–
4. Large problem priming effects	−.314**	−.058	−.049	–	–	−.236	−.187	−.013	–	–	−.475**	.283	−.167	–	–	−.316	.244	.120	–	–
5. Arithmetic fluency scores	−.448***	−.048	.232*	.213*	–	−.477**	.055	.200	.192	–	−.402	.164	.446**	.295	–	−.026	−.164	.448*	.155	–

* p < .05.
** p < .01.
*** p < .001.

Concerning large problems and the full sample of children, there was a positive correlation between arithmetic fluency scores and size effects (r = .232, p = .02), which remained significant after controlling for the age of children (r = .348, p < .001). To understand this correlation, it must be recalled that the slopes for large problems were negative in each age group. Therefore, higher children's math skills were related to more positive or less negative slopes. This result was observed in the two groups of older children (r = .446, p = .01 and r = .448, p = .03 for 10- to 12-year-olds and 13- to 15-year-olds, respectively), and the correlations remained significant after controlling for the age of children (r = .430, p = .01 and r = .450, p = .04 for 10- to 12-year-olds and 13- to 15-year-olds, respectively).

Moreover, in the full sample of children, there was a positive correlation between arithmetic fluency scores and priming effects (r = .213, p = .04), which remained marginally significant after the age of children was controlled for, (r = .181, p = .08).

Control analyses

Experiment 2 was always presented before Experiment 1. To evaluate whether this fixed order contaminated our crucial results on size effects, we conducted a series of t tests comparing size effects in each age group for small problems and large problems. Of the six t tests conducted, only one was significant (i.e., −44 ms for large problems in 8- to 9-year-olds in Experiment 2 and −150 ms in Experiment 1 [priming], $t(35) = -2.12$, p = .04), which ensures that the experiment fixed order of presentation had at worse only a minimal effect on our results.

General discussion

In this study, we have conjointly studied priming and size effects in three populations of children (8–9, 10–12, and 13–15 years of age) who were asked to solve addition problems. In this section, we address the results obtained on small problems before those obtained on larger problems.

For the group of older children, size effects for small additions involving operands from 2 to 4 were observed in association with priming effects of the "+" sign. However, this was not the case for the two groups of younger children, who also presented size effects for small problems but no evidence of priming effects. Therefore, the relatively large slopes of 282 ms in 8- to 9-year-olds and 155 ms in 10- to 12-year-olds must reflect the use of conscious reconstructive and counting strategies that cannot be primed in 150 ms by an arithmetic sign. In contrast, a smaller significant slope of less than 100 ms (i.e., 98 ms exactly) in older children may reveal the use of rapid, unconscious, and automatized counting procedures that can be primed by the anticipated presentation of the "+" sign. For the first time, therefore, we are able to quantify the mean size of the slope that can reflect the use of automated procedures at the group level.

However, it is probable that at an individual level the slope reflecting automatization is below 98 ms because some 13- to 15-year-old children might present a delay in the automatization process and, therefore, might present larger slopes, increasing the value of the average slope. As an illustration, an inspection of our data at an individual level reveals that 1 child in this age group presented a size effect of 637 ms. This shows that reasoning on averages over age group of children always needs to be done with care, mainly because of the variability in children's behaviors (Dewi & Thevenot, 2022; Siegler, 1987). Still, reasoning at the group level provides a reference value above which children could potentially be identified as presenting difficulties or delays in their procedure automatization (Bagnoud et al., 2021c). The negative correlation between arithmetic fluency scores and the size of the slope that we observe in the current research also attests that the size of the slope can constitute a valuable indicator of children's arithmetic difficulties. Noticeably, this correlation and all the others involving children's arithmetic fluency were still significant after controlling for the age of children. This means that children's level of expertise in arithmetic was specifically associated with the size of the slope independently of general cognitive maturation. Concerning the correlation between arithmetic fluency and the size of the slope that we have just mentioned, however, it should be noted that whereas this correlation was observed in our full sample of children as well as in our two younger groups of children, it was not significant in the group of 13- to 15-year-olds. A first explanation could be that the number of children involved in this group was too low compared with the other groups to reveal the correlation (n = 23 compared with 36 and 39 for the other two groups). A second explanation is that once the counting procedure is automatized, it no longer depends on arithmetic fluency because a threshold in arithmetic skills has been reached. This explanation will need to be investigated on larger samples of children because an argument against it is that automaticity is probably a continuous process rather than an on–off process (e.g., Logan, 1997). Moreover, we know that the size of the slope in adults is much lower (20–50 ms; Barrouillet & Thevenot, 2013; Uittenhove et al., 2016) than in this group of children (98 ms), which suggests that 13- to 15-year-olds could reach a higher degree of procedure automaticity later during their development (Jeon et al., 2019). However, we have already commented that an inspection of our data at an individual level shows that some children are still using costly conscious procedures (i.e., a slope of 637 ms for 1 child). Therefore, future research with more numerous 13- to 15-year-olds will be needed to determine whether a subpopulation of children, or only some exceptions, can be identified as lagging behind in terms of automatization.

If we envision that a minority of children have not reached automatization at 13 to 15 years of age, we also need to consider that a minority of 10- to 12-year-olds have already reached automatization of procedures for small problems. This assumption is supported by the negative correlation observed between size and priming effects in this age group, which could reflect the fact that some children who present a priming effect also present smaller slopes because they are more efficient. Still, it must be noted that this correlation between priming and size effects was not observed in Poletti et al. (2021) and, therefore, will need to be replicated before it can be discussed further.

Regardless of small subgroups of children who could be behind or ahead concerning the automatization process of counting procedures, it appears that counting automaticity occurs relatively late during development. Whereas automatization in reading develops during the course of the first year of instruction, typically in Grade 1 (Megherbi et al., 2018), the current results as well as previous results in the literature (Díaz-Barriga Yáñez et al., 2020; Mathieu et al., 2018a; Poletti et al., 2021) did not reveal signs of counting automaticity before Grade 6 or 7. It must be noted, however, that full automaticity in reading is not yet fully reached even after 4 years of instruction (Froyen et al., 2009). This suggests that it is a long road to automaticity whatever the cognitive domain considered. Still, it is plausible that full automaticity of counting procedures occurs later during development than reading automaticity because reading is a more frequent activity than counting both at school and in everyday life. Noticeably, the instructions in counting and arithmetic activities in schoolbooks are often read by children, and arithmetic exercises are often embedded in word problems (e.g., Verschaffel et al., 1999). The reverse, however, is obviously not true because there is no reason for reading activities to be connected to mathematical contents. A difference in the amount of practice between mathematical and reading activities (e.g., Eurydice, 2022), therefore, might be responsible for late automaticity of counting procedures compared with reading.

For larger problems with a sum from 7 to 10 involving one operand greater than 4, priming effects were not observed in any of the age groups. Moreover, and consistent with Uittenhove et al. (2016) in adults and Bagnoud et al. (2021a) in children and adults, no size effect was observed for these problems. This absence of variation in solution times (i.e., plateau) and the absence of priming effect of the "+" sign support the idea that the answers to such problems are retrieved from memory by children and adults. Note that, as already explained in our Introduction, considering size and priming effects conjointly allows us to resolve a serious difficulty of interpretation concerning the absence of priming effect. Indeed, without considering size effects, this absence can be interpreted either as the use of retrieval or as the use of conscious algorithmic procedures. Because the use of conscious algorithmic procedures necessarily results in a significant slope associating solution times with the size of the problem, a configuration where the absence of priming effect is observed with a lack of size effect, therefore, can be unambiguously interpreted as the use of retrieval. In contrast, and as already explained above in the discussion about small problems, a lack of priming associated with significant slopes must be interpreted by the use of conscious algorithmic procedures.

Interestingly, the absence of size effect for larger problems was also observed by Bagnoud et al. (2021a) already in children aged 6 to 7 years (Grade 1) for problems with sums to 8, 9, and 10. However, contrary to what is observed in older children and adults, sum to 7 problems were still subjected to size effects in this young population. Therefore, and contrary to what is expected following most retrieval models (e.g., Ashcraft, 1992; Siegler & Shrager, 1984), it seems that for problems with a sum up to 10, answers to larger problems are retrieved first during development (Compton & Logan, 1991; Dewi, Bagnoud, & Thevenot, 2021a, 2021b; Logan & Klapp, 1991; Thevenot et al., 2020). As stated above, whereas this last assumption cannot be accounted for by most retrieval models, there is one exception. Indeed, Logan, one of the most eminent retrieval theorists, described and explained retrieval of larger problems before smaller ones within a horse race model. Because for solving such problems the number of steps is too numerous to be conducted efficiently and rapidly, individuals would deliberately memorize their sums (Logan & Klapp, 1991). As is developed below, we have the same interpretation, but the difference between our conception and Logan's conception is that we do not assume that the response to small problems will eventually be retrieved later during development.

The conclusion that the answers to large problems are retrieved before the answers to smaller problems during development is also supported by the results obtained on large sum to 7 and 8 problems. In 8- to 9-year-old children, an increase in solution times between large problems with sums to 7 and 8 is observable. In contrast, there is no variation between problems with sums to 8 and 9, but there is a sharp decrease in solution times for sum to 10 problems. The special status of sum to 10 problems has already been theoreticized by retrieval model proponents (Campbell, 1995; Chen & Campbell, 2018). Sum to 10 problems would indeed be particularly salient and accessible due to the base 10 system. This special status of sum to 10 additions, which is evidenced by particularly short solution times in our study, is actually observable in each age group. In 10- to 12-year-old children, the difference in solution times between sum to 7 problems and sum to 8 problems, which was observable in younger children, has disappeared; therefore, the plateau in solution times is related to problems with sums from 7 to 9. In older children, the decrease in solution times starts from sum to 8 problems. These patterns of results can reflect a mix between procedural and retrieval strategies for large problems with an increase of retrieval reliance over development, starting with larger problems.

Such increasing reliance on retrieval with expertise, starting with the largest problems, would result in flatter slopes in experts than in novices. This is probably the explanation for the positive correlations we obtained between arithmetic fluency scores and the size of the slopes. Because for large problems the slopes that we observe are negative, this correlation shows that an increase in math fluency is associated with a decrease in the slope negativity. To confirm this interpretation, we compared the slopes of the 45 children with higher math skills (mean score of 78) with the slopes of the 45 children with lower skills (mean score of 40). To obtain clearly distinct groups, we removed the scores of 8 children around the median. In both groups, the slopes were negative. However, less negative slopes (−28 ms) were observed in the group of children with higher skills compared with the group of children with lower skills (−70 ms).

Following the logic that problems with the largest sum are retrieved first during development, it could be envisioned that the answers to small problems involving operands from 2 to 4 will eventually be retrieved from memory after 13 to 15 years of age. Nevertheless, there are two main arguments against this assumption. First, it could not explain why in older children small addition problems benefit from the anticipated presentation of the "+" sign, whereas larger problems do not. Second, it is well established that significant and positive size effects are observable for small problems during adulthood, that is, in a mature cognitive system (e.g., Barrouillet & Thevenot, 2013; LeFevre et al., 1996). These size effects were classically interpreted as variations in the speed of retrieval because of interference effects (Zbrodoff & Logan, 2005), but such explanations are now challenged by the observations that solution times can remain stable or even decrease, whereas the size of the problem increases. Indeed, if size effects in retrieval models are explained by the fact that larger problems suffer from more interference than smaller ones, a monotonic increase as a function of the size of the problem is expected.

A last correlation that we obtained, and that we have not discussed yet, is the one between priming effect and arithmetic fluency for large problems in our full sample of children. At first glance, the interpretation of this correlation appears to be difficult because priming effects are not observed at the group level for large problems. However, this correlation indicates that fluent children could present priming effects or, in other words, could use automated procedures, even when the additions involve operands outside the range of subitizing. This would challenge the automated counting theory. Nevertheless, even though subitizing is often viewed as limited to 4 objects (e.g., Revkin et al., 2008;

Simon & Vaishnavi, 1996), several studies suggest that the phenomenon could be extended to 5 objects (e.g., Starkey & Cooper, 1995; Wolters et al., 1987). Therefore, it is possible that some children in our sample have the ability to subitize up to 5 objects and, therefore, could have developed automated counting procedure involving operands up to 5. To examine this possibility, we split up our material of large problems with additions involving a 5, on the one hand (i.e., 5 + 2, 5 + 3, 5 + 4, and their inverse combinations), and problems with operands greater than 5, on the other (i.e., 6 + 2, 6 + 3, 6 + 4, 7 + 2, 7 + 3, 8 + 2, and their inverse combinations). Strikingly, we found that the correlation between arithmetic fluency and priming effects was limited to problems with operands up to 5 (r = .309, p = .002 vs. r = .062, p = .543 for problems with operands greater than 5). Therefore, our interpretation that some children in our sample have automatized counting procedures for a range of problems involving operands from 2 to 5 is plausible. However, this result will need to be replicated and tested directly before we can discuss it further and consider its implications for the automated counting theory.

All in all, our set of data is best explained by the idea that development toward expertise in solving small problems consists in an acceleration of one-by-one counting procedures until automatization. As was well described and summarized by Logan (1997), the main characteristics of automatized procedures are quickness, effortlessness, autonomy, and unconsciousness. It explains why, when solving small problems, experts are only aware of the final product of the procedure (i.e., the answer) and not of the procedure itself (i.e., counting steps). More precisely, when the time required to make a step is relatively long, the procedure is conscious and can be verbally reported by solvers. In contrast, very quick steps from one number to another (taking less than 100 ms, as reported in the current article) cannot reach consciousness provided that the number of steps to be performed is not too large. Indeed, to be adaptative, such automatized counting procedures are necessarily restricted to small quantities. Our results suggest that outside the range of subitizing, or the range of elements that can be apprehended in a single snapshot (Cowan, 2001), automatized procedures could be too slow and effortful to be efficient. In this case, retrieval of the answer from long-term memory could be the process maximizing performance. Therefore, our results could be explained by a horse race model (Logan & Cowan, 1984) between automatized counting procedures and retrieval in which procedures could win over retrieval when both operands are "subitizable." Mathieu et al. (2016) suggested that solving the problem by successive one-by-one increments could be done along a mental number line (see also Chouteau et al., 2021). Moves greater than four steps could not be possible without costly attentional shifts that would make the procedure inefficient and, therefore, would make retrieval the best strategy. As is well illustrated in the current article (Fig. 2), large problems solved mainly by retrieval are solved faster than small problems with a sum of 7 in 8- to 9-year-old children who have not yet automatized the counting procedures. However, this difference in solution times decreases until disappearance as children become expert and the counting procedures evolve toward automatization. At this point, automatized counting procedures can be faster than retrieval and win the race.

Stated differently and more generally, the overall pattern of results that we observed in this study could stem from the competition between procedures and retrieval for each problem. When the procedure is automatized, it virtually systematically wins the race; when it is not, retrieval wins the race, given that the association between the problem operands and the problem answer has already been stored in long-term memory. If this memory association has not been done yet, individuals need to rely on conscious and costly reconstructive strategies (Logan, 1988a, 1988b; Logan & Klapp, 1991).

In addition to their rapidity of execution, automated counting procedures present two other main advantages over retrieval. First, the addition facts represented in retrieval networks often interfere with each other, which is sometimes problematic and partly explains why individuals experience difficulties in learning and recalling them (e.g., De Visscher & Noël, 2014a, 2014b). Counting procedures are immune from such interference, and the automated counting process is guaranteed to run successfully to completion once it is launched (i.e., autonomy). Second, whereas procedures are transferable to new problems, this is not supposed to be the case for memorized facts (Campbell et al., 2016; Dewi & Thevenot, 2022; Logan & Klapp, 1991). Therefore, counting procedures are a more powerful tool than stored facts for generalization of learning (VanLehn, 1996).

A last point deserving discussion is the fact that the lack of priming effect for large problems in expert children could seem at odds with the observations of Fayol and Thevenot (2012) or Thevenot et al. (2020). Indeed, in these last studies, when problems are more roughly considered depending on their sums (i.e., ≤10 vs. ≥10) rather than depending on the size of the operands, priming effects for large problems are observed. Nevertheless, there is no contradiction between the different studies. All researchers in the domain of numerical cognition, whatever the theory they defend, agree that the large problems studied by Fayol and Thevenot are often solved by expert solvers via reconstructive strategies (e.g., Campbell & Timm, 2000; LeFevre et al., 1996; Thevenot et al., 2007). It is then plausible that automatized counting procedures are used to compute intermediate sums during the process of decomposition (e.g., 4 + 7 = 4 + 2 + 5; Cheng, 2012).

All in all, the results of this research have allowed us to put forward a new model of arithmetic skill development in which counting procedures limited to small problems can win in efficiency over retrieval. This model is completely original and even iconoclast because most classical models in the literature argue that during development retrieval will become the dominant strategy for all problems because it is faster than any reconstructive strategies (Ashcraft & Fierman, 1982; Siegler & Shrager, 1984). As claimed here, a conjoint examination of priming and size effects for addition problems does not support this assumption, at least not for problems constructed with operands within the subitizing range. Given that this range varies between individuals and that it can sometimes reach 5 items (Leibovich-Raveh et al., 2018), therefore, it is possible that some large problems in our study involving operands up to 5 are solved by some efficient children with automatized procedures. This could be the reason why, at the full sample level, a correlation between math skills and size of priming effect is observed even for large problems in the current research. This line of reasoning definitely needs to be developed in future experiments examining conjointly size and priming effects as well as individual subitizing ranges.

Before concluding, two limitations of the current research need to be mentioned. First, and as already evoked, the relatively small samples of participants in each of the age groups under study make it difficult to draw strong conclusions from the results of our correlational analyses. The relationships among priming effects, size effects, and individuals' arithmetic fluency will need to be addressed in future studies with larger samples of participants. Second, and maybe also because of a limited number of participants, ambiguous statistical results were obtained for the group of older children concerning priming effects. Whereas a priming effect of the "+" sign was found for small problems (i.e., shorter solution times in the negative SOA condition than in the null one) but not for larger problems, the interaction between SOA and problem size failed to reach significance ($p = .17$). The former result supports the idea that automated counting procedures are activated and used by 13- to 15-year-olds for small problems but not for large ones, but the latter result rather supports the position that both small and large problems are subjected to sign priming effects (given that the main effect of SOA was significant for this age group). However, this second possibility is strongly questioned by the fact that there was no increase in solution times for large problems depending on the size of problems. Counting or algorithmic procedures should indeed take longer as the size of the problems increases. Still, as evoked earlier, the possibility that automated counting procedures are possibly used beyond the range of very small problems (i.e., both operands up to 4) by individuals with particularly high working memory or attention capacities will be worth investigating in the future.

To conclude, the results of the current research question the existence of size effects systematically linking the size of simple addition problems to their solution times. Indeed, a precise examination of solution times depending on categories of problems established on theoretical elements (i.e., operands within or beyond the subitizing range) revealed an increase of solution times depending on the sum of small problems followed by a decrease in solution times for larger problems. Relating these size effects to the results obtained in a priming task of the "+" sign strongly suggests that, contrary to what is put forward by most retrieval models, expert solvers solve small addition problems by automatized counting procedures, whereas they retrieve the results of larger problems from memory.

Data availability

The link to the data is provided in the manuscript

Acknowledgments

This work was funded by the Swiss National Scientific Foundation (SNSF-100014L_182174) and the French National Research Agency (ANR-18-CE93-0001) to Catherine Thevenot and Jérôme Prado.

References

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition, 44*(1–2), 75–106. https://doi.org/10.1016/0010-0277(92)90051-i.

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. https://doi.org/10.1037/0278-7393.4.5.527.

Ashcraft, M. H., & Christy, K. S. (1995). The frequency of arithmetic facts in elementary texts: Addition and multiplication in Grades 1–6. *Journal for Research in Mathematics Education, 26*, 396–421. https://doi.org/10.2307/749430.

Ashcraft, M. H., & Fierman, B. A. (1982). Mental addition in third, fourth, and sixth graders. *Journal of Experimental Child Psychology, 33*, 216–234. https://doi.org/10.1016/0022-0965(82)90017-0.

Bagnoud, J., Dewi, J., Castel, C., Mathieu, R., & Thevenot, C. (2021a). Developmental changes in size effects for tie and non-tie addition problems in 6- to 12-year-old children and adults. *Journal of Experimental Child Psychology, 201*. https://doi.org/10.1016/j.jecp.2020.104987 104987.

Bagnoud, J., Dewi, J., & Thevenot, C. (2021b). Differences in event-related potential (ERP) responses to small tie, non-tie and 1-problems in addition and multiplication. *Neuropsychologia, 153*, 107771. https://doi.org/10.1016/j.neuropsychologia.2021.107771.

Bagnoud, J., Mathieu, R., Dewi, J., Masson, S., Gonzalez-Monge, S., Kasikci, Z., & Thevenot, C. (2021c). An investigation of the possible causes of arithmetic difficulties in children with dyscalculia. *Topics in Cognitive Psychology, 121*, 217–237. https://doi.org/10.3917/anpsy1.213.0217.

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. (1985). Mastery of basic number combinations: Internalization of relationships or facts? *Journal for Research in Mathematics Education, 16*(2), 83–98. https://doi.org/10.2307/748366.

Barrouillet, P., & Thevenot, C. (2013). On the problem size effect in small additions: Can we really discard any counting-based account? *Cognition, 128*, 35–44. https://doi.org/10.1016/j.cognition.2013.02.018.

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

Campbell, J. I., Chen, Y., Allen, K., & Beech, L. (2016). Transfer of training in alphabet arithmetic. *Memory & Cognition, 44*(8), 1288–1300. https://doi.org/10.3758/s13421-016-0631-x.

Campbell, J. I. D., & Thompson, V. A. (2012). Retrieval-induced forgetting of arithmetic facts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(1), 118–129. https://doi.org/10.1037/a0025056.

Campbell, J. I. D., & Timm, J. (2000). Adults' strategy choices for simple addition: Effects of retrieval interference. *Psychonomic Bulletin & Review, 7*, 692–699. https://doi.org/10.3758/BF03213008.

Carpenter, T. P., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education, 15*(3), 179–202. https://doi.org/10.5951/jresematheduc.15.3.0179.

Chen, Y., & Campbell, J. I. D. (2018). "Compacted" procedures for adults' simple addition: A review and critique of the evidence. *Psychonomic Bulletin & Review, 25*, 739–753. https://doi.org/10.3758/s13423-017-1328-2.

Chen, Y., & Campbell, J. I. D. (2019). Evidence for a retrieval over a sum counting theory of adults' simple addition. *Canadian Journal of Experimental Psychology, 73*(2), 69–78. https://doi.org/10.1037/cep0000171.

Cheng, Z. (2012). Teaching young children decomposition strategies to solve addition problems: An experimental study. *Journal of Mathematical Behavior, 31*, 29–47. https://doi.org/10.1016/j.jmathb.2011.09.002.

Chouteau, S., Mazens, K., Thevenot, C., Dewi, J., & Lemaire, B. (2021). Computational modeling of basic mental addition: A new approach based on ultra-fast and automated counting procedures along a mental line. In *Proceedings of the annual meeting of the Cognitive Science Society* (Vol. 43, pp. 2010–2016). Cognitive Science Society.

Compton, B. J., & Logan, G. D. (1991). The transition from algorithm to retrieval in memory-based theories of automaticity. *Memory & Cognition, 19*(2), 151–158. https://doi.org/10.3758/BF03197111.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922.

De Visscher, A., & Noël, M. P. (2014). The detrimental effect of interference in multiplication facts storing: Typical development and individual differences. *Journal of Experimental Psychology: General, 143*(6), 2380–2400. https://doi.org/10.1037/xge0000029.

De Visscher, A., & Noël, M.-P. (2014). Arithmetic facts storage deficit: The hypersensitivity-to-interference in memory hypothesis. *Developmental Science, 17*, 434–442. https://doi.org/10.1111/desc.12135.

Dewi, J., Bagnoud, J., & Thevenot, C. (2021a). Automatization through practice: The opportunistic-stopping phenomenon called into question. *Cognitive Science, 45*(12). https://doi.org/10.1111/cogs.13074. Article e13074.

Dewi, J., Bagnoud, J., & Thevenot, C. (2021b). Do production and verification tasks in arithmetic rely on the same cognitive mechanisms? A test using alphabet arithmetic. *Quarterly Journal of Experimental Psychology, 74*(12), 2182–2192. https://doi.org/10.1177/17470218211022635.

Dewi, J., & Thevenot, C. (2022). Individual differences in the evolution of counting: Evidence from transfer in alphabet arithmetic. *Experimental Psychology, 69*(2), 75–82. https://doi.org/10.1027/1618-3169/a000546.

Díaz-Barriga Yáñez, A., Couderc, A., Longo, L., Merchie, A., Chesnokova, H., Langlois, E., Thevenot, C., & Prado, J. (2020). Learning to run the number line: The development of attentional shifts using single-digit arithmetic. *Annals of the New York Academy of Sciences, 1477*(1), 79–90. https://doi.org/10.1111/nyas.14464.

Dowker, A. (2015). Individual differences in arithmetical abilities: The componential nature of arithmetic. In R. C. Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 878–894). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642342.013.034.

Eurydice (2022). *Primary education: Teaching and learning in primary education.* https://eurydice.eacea.ec.europa.eu/national-education-systems/france/teaching-and-learning-primary-education

Evans, T. M., & Ullman, M. T. (2016). An extension of the procedural deficit hypothesis from developmental language disorders to mathematical disability. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.01318. Article 1318.

Fayol, M., & Thevenot, C. (2012). The use of procedural knowledge in simple addition and subtraction problems. *Cognition, 123*, 392–403. https://doi.org/10.1016/j.cognition.2012.02.008.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*, 116–124. https://doi.org/10.3758/BF03195503.

Froyen, D. J., Bonte, M. L., van Atteveldt, N., & Blomert, L. (2009). The long road to automation: Neurocognitive development of letter–speech sound processing. *Journal of Cognitive Neuroscience, 21*(3), 567–580. https://doi.org/10.1162/jocn.2009.21061.

Grabner, R. H., Brunner, C., Lorenz, V., Vogel, S. E., & De Smedt, B. (2022). Fact retrieval or compacted counting in arithmetic—A neurophysiological investigation of two hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(2), 199–212. https://doi.org/10.1037/xlm0000982.

Groen, G. J., & Parkman, J. M. (1972). A chronometric analysis of simple addition. *Psychological Review, 79*, 329–343. https://doi.org/10.1037/h0032950.

Jeon, H. A., Kuhl, U., & Friederici, A. D. (2019). Mathematical expertise modulates the architecture of dorsal and cortico-thalamic white matter tracts. *Scientific Reports, 9*(1). https://doi.org/10.1038/s41598-019-43400-6. Article 6825.

Kaye, D. B., Post, T. A., Hall, V. C., & Dineen, J. T. (1986). Emergence of information-retrieval strategies in numerical cognition: A developmental study. *Cognition and Instruction, 3*, 127–150. https://doi.org/10.1207/s1532690xci0302_3.

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. https://doi.org/10.1037/0278-7393.22.1.216.

Leibovich-Raveh, T., Lewis, D. J., Kadhim, S. A. R., & Ansari, D. (2018). A new method for calculating individual subitizing ranges. *Journal of Numerical Cognition, 4*, 429–447. https://doi.org/10.5964/jnc.v4i2.74.

Logan, G. D. (1988a). Automaticity, resources, and memory: Theoretical controversies and practical implications. *Human Factors, 30*(5), 583–598. https://doi.org/10.1177/001872088803000504.

Logan, G. D. (1988b). Toward an instance theory of automatization. *Psychological Review, 95*, 492–527. https://doi.org/10.1037/0033-295X.95.4.492.

Logan, G. D. (1997). Automaticity and reading: Perspectives from the instance theory of automatization. *Reading & Writing Quarterly, 13*, 123–146. https://doi.org/10.1080/1057356970130203.

Logan, G. D. (2018). Automatic control: How experts act without thinking. *Psychological Review, 125*(4), 453–485. https://doi.org/10.1037/rev0000100.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review, 91*, 295–327. https://doi.org/10.1037/0033-295X.91.3.295.

Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(2), 179–195. https://doi.org/10.1037/0278-7393.17.2.179.

Mathieu, R., Epinat-Duclos, J., Léone, J., Fayol, M., Thevenot, C., & Prado, J. (2018a). Hippocampal spatial mechanisms scaffold the development of arithmetic symbol processing in children. *Developmental Cognitive Neuroscience, 30*, 324–332. https://doi.org/10.1016/j.dcn.2017.06.001.

Mathieu, R., Epinat-Duclos, J., Sigovan, M., Breton, A., Cheylus, A., Fayol, M., ... Prado, J. (2018b). What's behind a "+" sign? Perceiving an arithmetic operator recruits brain circuits for spatial orienting. *Cerebral Cortex, 5*, 1673–1684. https://doi.org/10.1093/cercor/bhx064.

Mathieu, R., Gourjon, A., Couderc, A., Thevenot, C., & Prado, J. (2016). Running the number line: Operators elicit horizontal shifts of attention during single-digit arithmetic. *Cognition, 146*, 229–239. https://doi.org/10.1016/j.cognition.2015.10.002.

McDougle, S. D., Tsay, J. S., King, M., Pitt, B., Saban, W., Taylor, J. A., & Ivry, R. B. (2022). Continuous manipulation of mental representations is compromised in cerebellar degeneration. *Brain, 145*(12), 4246–4263. https://doi.org/10.1093/brain/awac072.

Megherbi, H., Elbro, C., Oakhill, J., Segui, J., & New, B. (2018). The emergence of automaticity in reading: Effects of orthographic depth and word decoding ability on an adjusted Stroop measure. *Journal of Experimental Child Psychology, 166*, 652–663. https://doi.org/10.1016/j.jecp.2017.09.016.

Pinheiro-Chagas, P., Dotan, D., Piazza, M., & Dehaene, S. (2017). Finger tracking reveals the covert stages of mental arithmetic. *Open Mind, 1*(1), 30–41. https://doi.org/10.1162/OPMI_a_00003.

Poletti, C., Perez, J. F., Houillon, J. C., Prado, J., & Thevenot, C. (2021). Priming effects of arithmetic signs in 10- to 15-year-old children. *British Journal of Developmental Psychology, 39*(3), 380–392. https://doi.org/10.1111/bjdp.12363.

Protopapas, A. (2007). CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods, 39*, 859–862. https://doi.org/10.3758/BF03192979.

Revkin, S. K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychological Science, 19*(6), 607–614. https://doi.org/10.1111/j.1467-9280.2008.02130.x.

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. https://doi.org/10.1080/09541440042000115.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General, 116*, 250–264. https://doi.org/10.1037/0096-3445.116.3.250.

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

Siegler, R. S., & Jenkins, E. (1989). *How children discover new strategies*. Lawrence Erlbaum.

Siegler, R. S., & Robinson, M. (1982). The development of numerical understanding. In H. W. Reese & P. Lipsitt (Eds.). *Advances in child development and behavior* (Vol. 16, pp. 241–312). Academic Press.

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). Lawrence Erlbaum.

Simon, T. J., & Vaishnavi, S. (1996). Subitizing and counting depend on different attentional mechanisms: Evidence from visual enumeration in afterimages. *Perception & Psychophysics, 58*, 915–926. https://doi.org/10.3758/BF03205493.

Starkey, P., & Cooper, R. G. (1995). The development of subitizing in young children. *British Journal of Developmental Psychology, 13*(4), 399–420. https://doi.org/10.1111/j.2044-835X.1995.tb00688.x.

Svenson, O. (1985). Memory retrieval of answers of simple additions as reflected in response latencies. *Acta Psychologica, 59*, 285–304. https://doi.org/10.1016/0001-6918(85)90048-4.

Thevenot, C., & Barrouillet, P. (2020). Are small additions solved by direct retrieval from memory or automated counting procedures? A rejoinder to Chen and Campbell (2018). *Psychonomic Bulletin & Review, 27*(6), 1416–1418. https://doi.org/10.3758/s13423-020-01818-4.

Thevenot, C., Barrouillet, P., Castel, C., & Uittenhove, K. (2016). Ten-year-old children strategies in mental addition: A counting model account. *Cognition, 146*, 48–57. https://doi.org/10.1016/j.cognition.2015.09.003.

Thevenot, C., Castel, C., Danjon, J., & Fayol, M. (2015). Identifying strategies in arithmetic with the operand recognition paradigm: A matter of switch-cost? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 541–552. https://doi.org/10.1037/a0038120.

Thevenot, C., Dewi, J., Bagnoud, J., Uittenhove, K., & Castel, C. (2020a). Scrutinizing patterns of solution times in alphabet–arithmetic tasks favors counting over retrieval models. *Cognition, 200*. https://doi.org/10.1016/j.cognition.2020.104272 104272.

Thevenot, C., Dewi, J., Bagnoud, J., Wolfer, P., Fayol, M., & Castel, C. (2020b). The use of automated procedures by older adults with high arithmetic skills during addition problem solving. *Psychology & Aging, 35*, 411–420. https://doi.org/10.1037/pag0000431.

Thevenot, C., Fanget, M., & Fayol, M. (2007). Retrieval or non-retrieval strategies in mental addition? An operand-recognition paradigm. *Memory & Cognition, 35*(6), 1344–1352. https://doi.org/10.3758/BF03193606.

Uittenhove, K., Thevenot, C., & Barrouillet, P. (2016). Fast automated counting procedures in addition problem solving: When are they used and why are they mistaken for retrieval? *Cognition, 146*, 289–303. https://doi.org/10.1016/j.cognition.2015.10.008.

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology, 47*, 513–539. https://doi.org/10.1146/annurev.psych.47.1.513.

Verschaffel, L., De Corte, E. D., & Vierstraete, H. (1999). Upper elementary school pupils' difficulties in modeling and solving nonstandard additive word problems involving ordinal numbers. *Journal for Research in Mathematics Education, 30*, 265–285. https://doi.org/10.2307/749836.

Wolters, G., Van Kempen, H., & Wijlhuizen, G.-J. (1987). Quantification of small numbers of dots: Subitizing or pattern recognition? *American Journal of Psychology, 100*(2), 225–237. https://doi.org/10.2307/1422405.

Zbrodoff, N. J. (1995). Why is 9 + 7 harder than 2 + 3? Strength and interference as explanations of the problem-size effect. *Memory & Cognition, 23*, 689–700. https://doi.org/10.3758/BF03200922.

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). Psychology Press.