


**Scrutinizing Patterns of Solution Times in Alphabet-arithmetic Tasks
Favors Counting over Retrieval Models**


Catherine Thevenot^{1,*}, Jasinta D. M. Dewi^{1,*}, Jeanne Bagnoud¹,
Kim Uittenhove², and Caroline Castel²

¹University of Lausanne, Institute of Psychology, Switzerland

²University of Geneva, Department of Psychology, Switzerland

Author note

Catherine Thevenot  <https://orcid.org/0000-0002-4997-1882>

Jasinta Dewi  <https://orcid.org/0000-0003-4443-6634>

Jeanne Bagnoud  <https://orcid.org/0000-0002-3323-8467>

Kim Uittenhove  <https://orcid.org/0000-0001-5450-3875>

Kim Uittenhove is now at the Institute of Psychology of the University of Lausanne.

A part of the results has been already presented in Oxford, UK at the Mathematical Cognition and Learning Society Conference (2018) and in Granada, Spain at the International Meeting of the Psychonomic Society (2016).

This work was supported by Grant 100014_138185 from the Swiss National Foundation for Scientific Research to Catherine Thevenot and a grant from the Chuard-Schmid Foundation of the University of Lausanne to Catherine Thevenot. We would like to thank Marie-Pascale Noël for suggesting us the end-term effects as the explanation for the drop in solution times.

* These two authors have contributed equally to the study and should be recognized as co-first authors

Correspondence concerning this article should be addressed to Catherine Thevenot, University of Lausanne, SSP, Institute of Psychology, Géopolis Building, Room 4536, CH-1015 Lausanne, Switzerland. Email: catherine.thevenot@unil.ch

Abstract

According to associationist models, initial sequential processing of algorithmic steps is replaced through learning by single-step access to a memory instance. In an alphabet-arithmetic task where equations such as $C+3=F$ have to be verified, the shift from algorithmic procedures to retrieval would manifest in a transition from steep slopes relating solution times to addends at the beginning of learning to a flat function at the end (e.g., Logan & Klapp, 1991). Nevertheless, we argue that computation of the slopes at the end of training is biased by a systematic drop in solution times for the largest addend in the study set. In this paper, this drop is observed even when the longest training period in alphabet-arithmetic literature is doubled (Experiment 1) and even when the size of the largest addend is increased (Experiment 2). We demonstrate that this drop is partly due to end-term effects but remains observable even when end-term problems are not considered in the analyses. As Logan and Klapp suggested, we conclude that the drop is partly due to deliberate memorization of the problems with the largest addend. In contrast, departing from Logan and Klapp, we demonstrate that, when problems with the largest addend are excluded from the analyses, the possibility that counting is still used after learning cannot be discarded. This conclusion is reached because after this exclusion, the slopes were still significant. To conclude, our results advocate that practicing an algorithm leads to its acceleration and not to a shift from algorithmic procedures to retrieval.

Keywords: Numerical cognition; Arithmetic; Strategies; Counting; Automatization

Scrutinizing Patterns of Solution Times in Alphabet-arithmetic Tasks

Favors Counting over Retrieval Models

One of the fundamental mechanisms of learning is the possibility for basic cognitive systems to form associations between different elements in long-term memory. In humans, these associations are not only the result of perception but can also be produced by mental computation (Thorndike, 1911). These assumptions have guided the influential instance theory of automatization formulated by Logan (1988). According to this theory, any mental operation initially executed through an algorithm eventually leads to the construction of a trace associating the elements of the operation in memory. One additional trace is created each time the mental operation is executed. After repeated practice, the number of traces of the association is high enough for this association to be retrieved from long-term memory without any further recourse to the initial algorithm. For example, repeatedly solving a problem would eventually create an association between the question and the answer. Directly retrieving this association from long-term memory is viewed as the most efficient solving strategy because it is achievable in a single step (Logan, 1988).

This theoretical framework has found one of its most perfect fields of application in the domain of mental arithmetic and simple addition problem solving. In children and in adults, solution times increase with the size of the smaller operand involved in the problem. However, whereas the slope of the regression line is about 400 ms/increment in 6-year-old children, it is only of 20 ms/increment in adults (Groen & Parkman, 1972). The steep slope of 400 ms/increment undoubtedly reflects counting procedures in children but the authors considered that 20 ms/increment was too short to reflect a plausible counting mechanism in adults. Therefore, they concluded that adults usually retrieve the answers of simple additions from long-term memory but that the retrieval process fails randomly for 5% of the trials. This occasional failure would be responsible for the 20-ms/increment slope. In perfect coherence

with the associationist theory, and more precisely with the instance theory of automatization, the shift from counting to retrieval during the course of development has been interpreted as the result of repeated practice of counting procedures, which would ultimately lead to the association between the operands and the answers and to the creation of arithmetic facts in long-term memory (Geary, 1996; Logan, 1988; Logan & Klapp, 1991; Siegler & Jenkins, 1989; Siegler & Shipley, 1995; Siegler & Shrager, 1984). Retrieval of arithmetic facts is therefore considered as the dominant strategy used to solve simple addition problems in adults. Nowadays, increase in solution times is not only viewed as consequences of reversal to primitive methods in case of retrieval failures but also as the result of interference and frequency effects within a retrieval network (e.g., Ashcraft, 1982; 1992; Campbell, 1995; Campbell & Oliphant, 1992; Chen & Campbell, 2018; Siegler, 1996). More precisely, small problems would be solved more frequently than larger ones, which would result in higher memory access for small problems. Additionally, large problems share their sums with more problems than small ones, which would result in facilitated access to small problems suffering from less interference.

However, recent studies suggest that retrieval of arithmetic facts may not be the dominant strategy in adults and that automatized and unconscious procedural counting can also account for simple addition problem solving in expert adults. This challenging hypothesis was first formulated after it was shown that problem solving is facilitated when the arithmetic sign is presented 150 ms before the operands for simple additions but not for multiplications (Fayol & Thevenot, 2012). It was inferred from these results that abstract procedures were primed by the “+” sign and subsequently used to solve addition problems. Priming effects of the “+” sign are observed starting from the age of 12–13 years, which suggests that automatized addition procedures would emerge around this developmental point (Mathieu, Epinat-Duclos, Léone et al., 2018). Fayol and Thevenot as well as Barrouillet and Thevenot

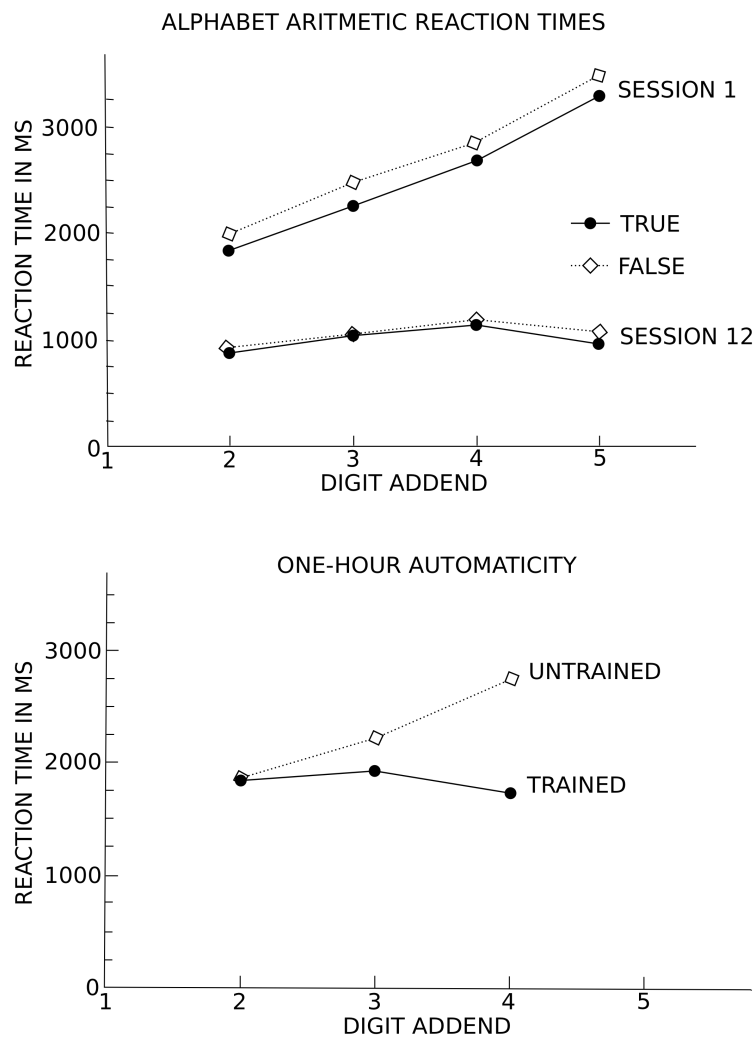
(2013) suggested that these procedures could correspond to quick scrolling of ordered numerical representations such as a number line or a verbal number sequence. This interpretation finds support in neuroimaging studies showing that multiplications are associated with brain regions involved in verbal retrieval, whereas single-digit additions are associated with brain regions involved in spatial attentional processing (Mathieu, Epinat-Duclos, Sigovan et al., 2018; Zhou et al., 2007). At a behavioral level, it has also been shown that addition problems are associated with spatial shifts of attention to the right hemifield (Liu et al., 2017; Masson et al., 2018; Zhu et al., 2018) whereas no spatial shifts of attention are observed for multiplication problems (Li et al., 2018; Mathieu et al., 2016). In fact, these different results could rehabilitate an idea that has been virtually abandoned more than 30 years ago by the research community and according to which arithmetic fluency is not necessarily achieved by arithmetic fact retrieval but can also be achieved by the automatization of rules and heuristics (Baroody, 1983; 1984; 1994; 2018).

However, the view that one of the most recurrent mental activities, namely basic arithmetic, does not always lead to associations between elements is necessarily at odds with the conception that memory retrieval is the process underlying automaticity after extensive practice. More precisely, it is in conflict with the conclusion of one of the most influential studies in the literature supporting the instance theory of automatization in which Logan and Klapp (1991) trained 8 participants on an alphabet-arithmetic task. Over a 12-day period, participants learnt to add the digits 2, 3, 4 and 5 either to the letters A to J or to the letters K to T. In the 13th session, participants were tested on the untrained set of the alphabet and the authors examined transfer of practice to new items. The results showed that participants improved their performance drastically from sessions 1 to 12. In the first session, solution times increased considerably with the magnitude of the addend. The function was clearly linear, with a slope of 486 ms/addend. By the 12th session, the function flattened substantially.

The slope was reduced to 45 ms/addend and the linear trend was no longer significant (see the top panel of Figure 1).

The authors concluded that Session 1 was dominated by counting algorithms because solution times increased linearly as a function of the addend. On the contrary, the flattened slope in Session 12 was thought to reflect memory retrieval. The authors acknowledged that their results could be attributable to an improvement in the speed of counting algorithms rather than a shift from counting to retrieval but were quick to dismiss this possibility, based on the argument that the slope on Session 12 averaged 45 ms/addend, which was viewed as too fast to be mediated by sequential access to the alphabet (e.g., $C + 3 = D, E, F$). However, drawing conclusions based on those 45 ms/addend could be questioned.

In fact, according to the authors, the slope of 45 ms/addend obtained at the end of training was the result of two different processes. Whereas the shift from procedure to retrieval was progressive across the 12 sessions for problems involving addends 2, 3 and 4, the shift was sudden for addend 5 problems between sessions 4 and 5 (see also Logan, 1988). More precisely, after only several sessions, solution times for addend 5 problems dropped noticeably and the solving process became faster than for problems involving addend 4. The decrease in solution times for addend 5 during session 12 is clear on the top panel of Figure 1. The authors explained the discontinuity in solution times for +5 problems by the fact that, for these problems and from session 4 onwards, their participants chose to use deliberate mnemonic encoding strategies. Logan and Klapp (1991) contrasted this deliberate memorization for addend 5 problems from “a relatively “automatic” encoding of the counting episode” (p. 183) for addends 2, 3 and 4 problems. Therefore, for the authors two different mechanisms allow for the shift of strategy from counting to retrieval.

Figure 1*Mean Solution Times as a Function of Addend*

Note. Top panel: Mean solution times for Sessions 1 (top lines) and 12 (bottom lines) in Logan and Klapp's (1991) Experiment 1 for true (circles, solid lines) and false (diamonds, dotted lines) equations. Bottom panel: Mean solution times for trained (circles, solid line) and untrained (diamonds, dotted line) items in Logan and Klapp's (1991) Experiment 2. Adapted from "Automatizing Alphabet Arithmetic: I. Is Extended Practice Necessary to Produce Automaticity?", by G. D. Logan and S. T. Klapp, 1991, *Journal of Experimental Psychology: Learning, Memory and Cognition*, 17, p. 182 (top panel) and p. 186 (bottom panel). Copyright 1991 by the American Psychological Association.

Nevertheless, the interpretation that the drop in solution times for +5 problems in Logan and Klapp's first experiment (1991) is due to the contrast between automatic encoding strategies and deliberate memorization is difficult to reconcile with the results of their second experiment. In this experiment, they asked their participants to study alphabet-arithmetic problems involving addends from 2 to 4 exclusively by deliberate memorization.

Nevertheless, a drop in solution times was still observed for the largest addend. Shorter solution times were indeed observed for deliberately memorized problems involving the highest addend 4 than for deliberately memorized problems with addend 3 (see the bottom panel of Figure 1). Thus, the drop in solution times cannot be attributable only to the fact that the problems with the highest addend are deliberately memorized. Therefore, another explanation of the discontinuity in solution times after extensive training in an alphabet-arithmetic task might be needed.

As a matter of fact, end-term effects constitute a potential candidate in order to explain that, after training, +5 problems are solved faster than +4 problems. End-term stimuli can be processed without having to pay attention to the whole problem because one of their components is systematically associated with the same response (Potts et al., 1978). For example, in a digit comparison task, the digit 9 is always the largest and the digit 1 is always the smallest of the pair of digits to be compared. Then, if the task requires the selection of the larger of two digits, 1 is never the appropriate choice but 9 is always the appropriate choice, irrespective of the other element in the pair (Moyer & Landauer, 1967). Therefore, when a terminal marker is found in an equation or a problem, the decision process terminates without the need to examine the other elements of the equation or problem (Foos et al., 1976). Short reaction times are therefore expected for stimuli corresponding to endpoints within the study set. More concretely in the alphabet-arithmetic task, participants who are trained with letters A to J and with addends from 1 to 5 can decide that $J + 5 = P$ is false because whenever P is

presented, the response is always “false”. There is therefore no need to examine the first part of the equation and the decision is taken very quickly. Note however, as it can be seen on the top panel of Figure 1, that the drop in solution times for +5 problems is also observed for problems associated with the true answer, for which partial processing is not possible before decision. Nevertheless, the answer representing the endpoint within the set of true problems also have a special status because it is presented only twice within the whole set of problems whereas the true answer for non-end-term problems is presented more than twice. $J + 5$ is associated with the true answer O and the only other occurrence of O is as a false answer (i.e., true answer + 1) to the problem $J + 4$. In contrast, the true answer for $C + 4$, which is G, appears as the true answer also to $E + 2$, $D + 3$ and $B + 5$ as well as a false answer to $B + 4$, $C + 3$, $C + 5$, $D + 2$, $D + 4$, $E + 3$, $F + 2$. Therefore, the answers to end-term problems suffer from less fan effect (Anderson & Reder, 1999) than the answers to non-end-term problems, whose multiple associations produce more interference. Stated differently, the answer to end-term problems is more salient because it is shared by a limited number of problems (Campbell et al., 2016). As a consequence, end-term problems should receive more attention and be processed quicker than less salient problems.

If such end-term effects explain why a drop in solution times is associated with the largest addend in alphabet-arithmetic tasks, including end-term problems in the estimation of the slopes relating solution times to the magnitude of the addends might lead to improper conclusions. It is apparent on the top panel of Figure 1 that solution times for addends 2, 3 and 4 follow a clear linear function. An estimation based on this Figure reveals that a slope of about 125 ms/addend is still observable after 12 practice sessions, whereas Logan and Klapp (1991) based their conclusions on a much flatter slope of 45 ms/addend when problems with addend 5 were included. Interestingly, the slope of 125 ms per letter corresponds to the overt or subvocal recitation speed reported by Landauer (1962). In this context, discarding the

hypothesis that fast counting is used by participants at the end of the alphabet-arithmetic training on the basis of the 45 ms/addend slope might be premature. On a more theoretical level, this would be crucial to ensure the credibility of the automatization of counting procedure model, according to which expertise in addition results from the acceleration of counting rather than a shift from counting to retrieval.

In order to shed light on this matter, we intended to ensure that the residual slope of about 125 ms/addend after 12 training sessions in Logan and Klapp's experiment (1991) was not due to an insufficient number of sessions for a possible shift to retrieval. Therefore, in a first experiment, we extended Logan and Klapp's study over 25 sessions instead of 12, with a total of 300 presentations of each problem instead of 144 in the original study. According to the authors, increasing the number of problem presentations also increase the degree of automaticity at the end of training (Logan & Klapp, Experiment 3). Moreover, in order to improve the reliability of our conclusions, 19 participants instead of 8 in the original study were involved in the training program. If the drop in solution times for addend 5 problems is still observed, we will analyze the results without end-term stimuli or, in other words, without solution times related to the equations $J + 5 = O$ and $T + 5 = Y$. We predict that end-term problems are responsible for the drop in solution times.

EXPERIMENT 1

Method

Participants. Nineteen students aged between 18 and 35 years were recruited at the University of Geneva. They received CHF 200 for their participation. In order to increase their motivation, participants were informed that a CHF 50 bonus would be awarded to those who obtained the best performance during the training phase. Informed consent was obtained for each of the participant.

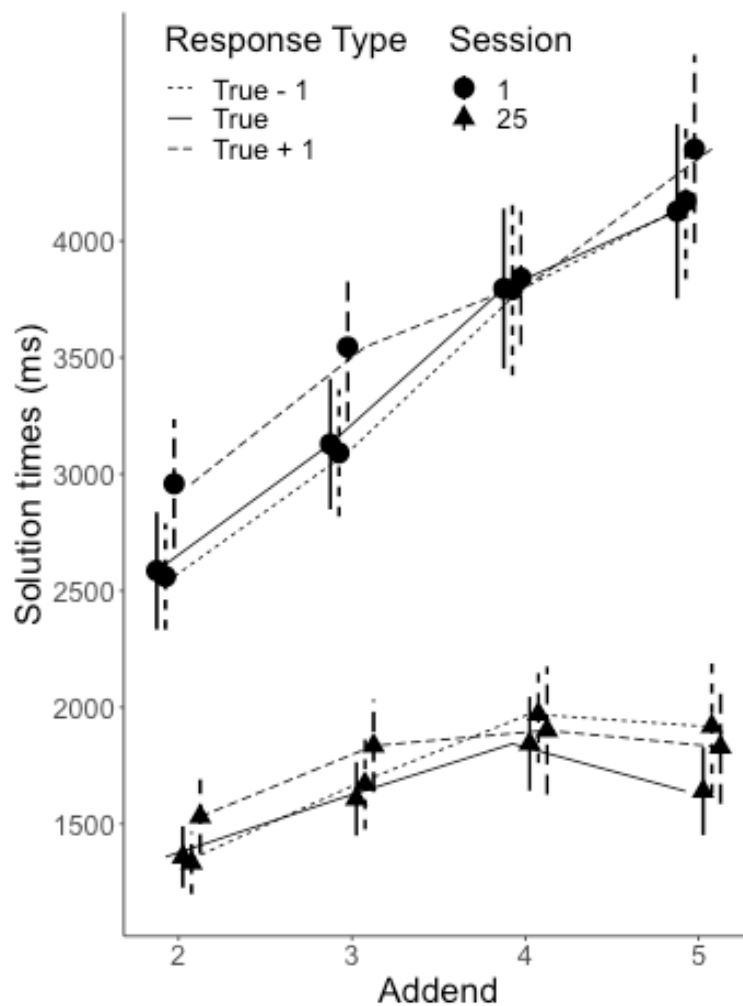
Material and procedure. The experiment was constructed as a training study, similar to the one designed by Logan and Klapp (1991). Participants were trained on an alphabet-arithmetic verification task wherein equations in which digits are added to letters have to be verified (e.g., $A + 2 = D$; FALSE). Half of the presented equations were associated with the correct letter (e.g., $A + 3 = D$) whereas the other half was associated with an incorrect answer (e.g., $C + 4 = H$). Half of the incorrect answers corresponded to the letter preceding the correct answer while the other half corresponded to the letter succeeding it. Participants were trained on equations with addends from 2 to 5, associated either with the first 10 letters of the alphabet (i.e., A to J) or with the next 10 letters (i.e., K to T). Participants were therefore trained on 40 problems and were randomly assigned to the first or second part of the alphabet.

A computerized version of the task was created using E-prime 2.0 software. Material was set up on participants' laptops for home training. The equations were presented horizontally in the center of the screen. Each trial began with a fixation point (*) presented for 500 ms, followed by the equation, which remained on the screen until the participant pressed the response key. Participants were required to press the "A" key when the presented equation was correct and the "L" key when it was incorrect. Then, the screen remained blank for 1500 ms until the onset of the next trial. Every combination of letters (A to J or K to T), addends (2 to 5), and response validity (TRUE or FALSE) was presented six times per session. Thus, every session involved 480 trials (i.e., 10 letters x 4 addends x 2 possible answers x 6 repetitions), which were divided into four blocks separated by a break. At the end of every session, the percentage of correct responses was displayed and participants had to note it down in a table. Twenty-five training sessions took place over 25 to 30 days because participants were allowed to have a one-day break during the week.

Results

Accuracy. Overall, the percentage of correct responses was of 98% ($SD = 2\%$), increasing from 96% in Session 1 to 99% in Session 25. A 2 (Letter Set: A to J vs. K to T) x 25 (Session: 1 to 25) x 4 (Addend: 2 to 5) mixed-design ANOVA was conducted on correct responses with Letter Set as a between measure. There was no effect of Letter Set and this variable did not interact with Session or Addend ($F_s < 1$). Therefore, we collapsed the two letter sets and conducted a 25 (Session: 1 to 25) x 4 (Addend: 2 to 5) repeated-measures ANOVA, which confirmed that the rates of correct responses increased across sessions ($F(24, 432) = 7.52, \eta_p^2 = .29, p < .001$) and revealed that they decreased as a function of addends ($F(3, 54) = 3.10, \eta_p^2 = .15, p = .03$). There was no interaction between Session and Addend ($F < 1$). Because error rates were low, the remainder of the result section will focus on solution times.

Solution times. We observed that not only were solution times modulated by the type of answer presented to participants (i.e., true or false) but also by the type of false answer (i.e., +1 or -1 from the correct answer, see Figure 2). Therefore, in order to draw conclusions from the cleanest results, our analyses will concern only the true equations correctly solved by participants. Before conducting the analyses, 1.3% of the solution times that could not correspond realistically to processing time of the whole problem (i.e., shorter than 300 ms) or that were too large to correspond to processing time only (i.e., longer than mean + 3 SD per participant and per session) were identified as outliers and were discarded from the analyses.

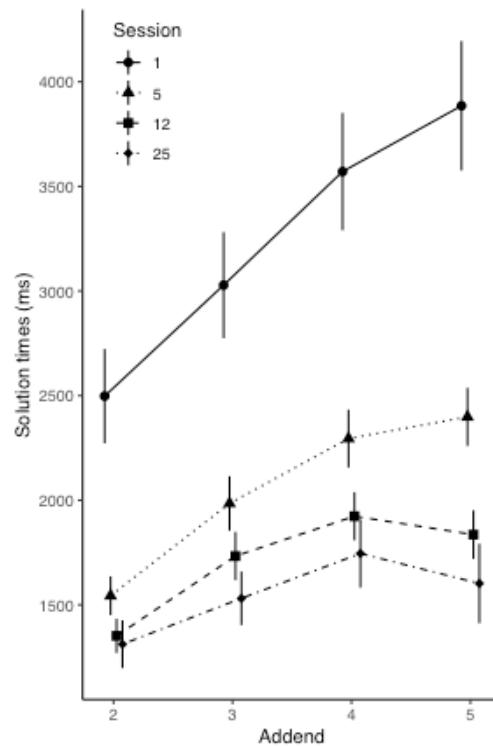
Figure 2*Mean Solution Times as a Function of Addend*

Note. Mean solution times as a function of addends for Sessions 1 (circles) and 25 (triangles) according to the nature of the presented answer, i.e. true (solid lines), true-1 (dotted lines) and true+1 (dashed lines) equations. Error bars represent standard errors.

A 2 (Letter Set: A to J vs. K to T) x 25 (Session: 1 to 25) x 4 (Addend: 2 to 5) mixed-design ANOVA with Letter Set as a between measure was conducted on solution times. There was no effect of Letter Set ($F(1, 17) = 1.28, p = .27$) or interactions between Letter Set and Session or Addend ($F_s < 1$). After collapsing the two letter sets, we conducted a 25 (Session: 1 to 25) x 4 (Addend: 2 to 5) repeated-measures ANOVA. The mean solution times across

sessions was of 1840 ms ($SD = 1000$ ms) and significantly decreased over sessions, from 3246 ms to 1547 ms ($F(24, 432) = 39.10, \eta_p^2 = .68, p < .001$). A main effect of Addend was also found ($F(3, 54) = 26.22, \eta_p^2 = .59, p < .001$), showing that solution times generally increased as a function of addends (i.e., 1461, 1821, 2074, and 2035 ms for Addend 2, 3, 4, and 5 respectively).

More important in our results was the interaction between Session and Addend ($F(72, 1296) = 8.94, \eta_p^2 = .33, p < .001$, see Figure 3) showing that the effect of Addend decreased across sessions. Nevertheless, when the slope of solution times as a function of addends, hereafter addend slope, was calculated for each participant and each session, it remained significant from the beginning to the end of training (Session 1: $M = 472$ ms/addend, $SE = 41, p < .001$; Session 12: $M = 163$ ms/addend, $SE = 42, p = .001$; Session 25: $M = 108$ ms/addend, $SE = 45, p = .03$). Whereas the difference in addend slope between Session 1 and Session 12 was significant ($t(18) = 4.92, p < .001$), this was not the case between Sessions 12 and 25, ($t(18) = 1.85, p = .08$).

Figure 3*Mean Solution Times as a Function of Addend*

Note. Mean solution times as a function of addends for Sessions 1 (circles, solid line), 5 (triangles, dotted line), 12 (squares, dashed line), and 25 (diamonds, dot-dashed line) for true trials. Error bars represent standard errors.

In order to determine whether solution times increased regularly with addends or in other words whether the effect of addend on solution times followed a linear trend, we carried out a series of contrast with Holm correction. In Session 1, the linear ($t(18) = 11.53, p < .001$) and quadratic ($t(18) = -2.57, p = .04$) trends were significant whereas in Session 25, the linear trend became marginal ($t(18) = 2.40, p = .055$) but the quadratic trend remained significant ($t(18) = -2.91, p = .03$).

In accordance with this quadratic trend and similar to the results of Logan and Klapp (1991), we found a discontinuity in solution times at +4 problems. Descriptively, solution times increased noticeably from addends 4 to 5 in the first session (+319 ms), but decreased in

the middle session (−89 ms), and even more in the last session (−144 ms). As planned and in order to investigate the possible explanation of this decrease by end-term effects, we removed J + 5 and T + 5 problems from the data set. Even in this case, +5 problems were solved faster than +4 problems during session 12 (−44 ms) and session 25 (−92 ms). When we recalculated the slopes without J + 5 and T + 5 problems, we found that they reached 180 ms/addend in session 12 (against 163 ms/addend when all problems were considered) and 128 ms/addend in session 25 (against 108 ms/addend when all problems were considered).

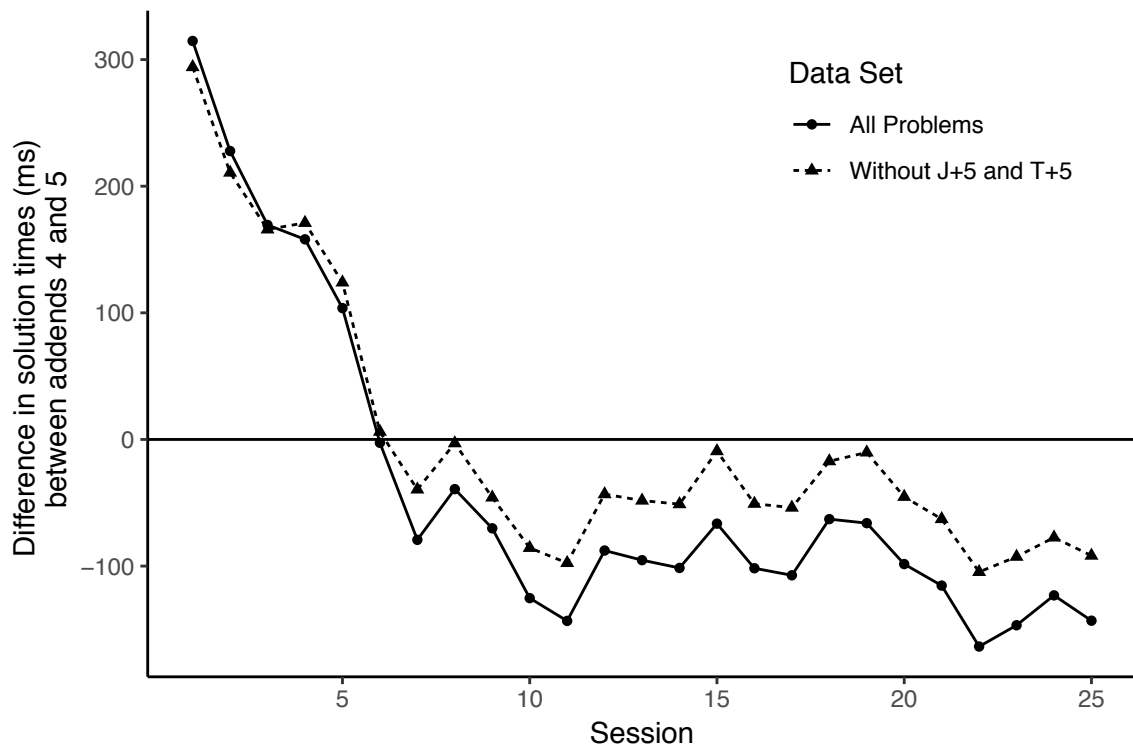
In order to further investigate the discontinuity in solution times at +4, we plotted the difference in solution times between +5 and +4 problems against sessions (see Figure 4 in which negative differences correspond to shorter solution times for +5 than for +4 problems). It can be seen that problems with addend 5 were solved faster than problems with addend 4 from session 6 until the end of the training and for every single session. This was true when we included all the problems as well as when we excluded the end-term problems. The increasing negative difference in solution times between addends 5 and 4 across sessions allows us to explain the increasing importance of the quadratic trend over the linear one.

Therefore and obviously, at least from session 6, +5 problems were not processed as the other problems. This is the reason why we recalculated the addend slopes from addend 2 to addend 4 and compared them to the slope when all addends were considered. In Session 1, the slope for addends 2 to 4 corresponded to 536 ms/addend and differed significantly from the slope of 472 ms/addend for addends 2 to 5 ($t(18) = 2.64, p = .02$). In Session 12, the slope for addends 2 to 4 corresponded to 286 ms/addend and differed significantly from the slope of 163 ms/addend for addends 2 to 5 ($t(18) = 2.75, p = .01$). Finally, in Session 25 the slope for addends 2 to 4 corresponded to 217 ms/addend and differed significantly from the slope of 108 ms/addend for addends 2 to 5 ($t(18) = 3.37, p = .003$). When +5 problems were not considered, the difference in addend slope between Session 1 and Session 12 was significant

($t(18) = 4.08, p < .001$) but this was not the case between Session 12 and 25, ($t(18) = 1.82, p = .09$).

Figure 4

Difference in Solution Times as a Function of Session



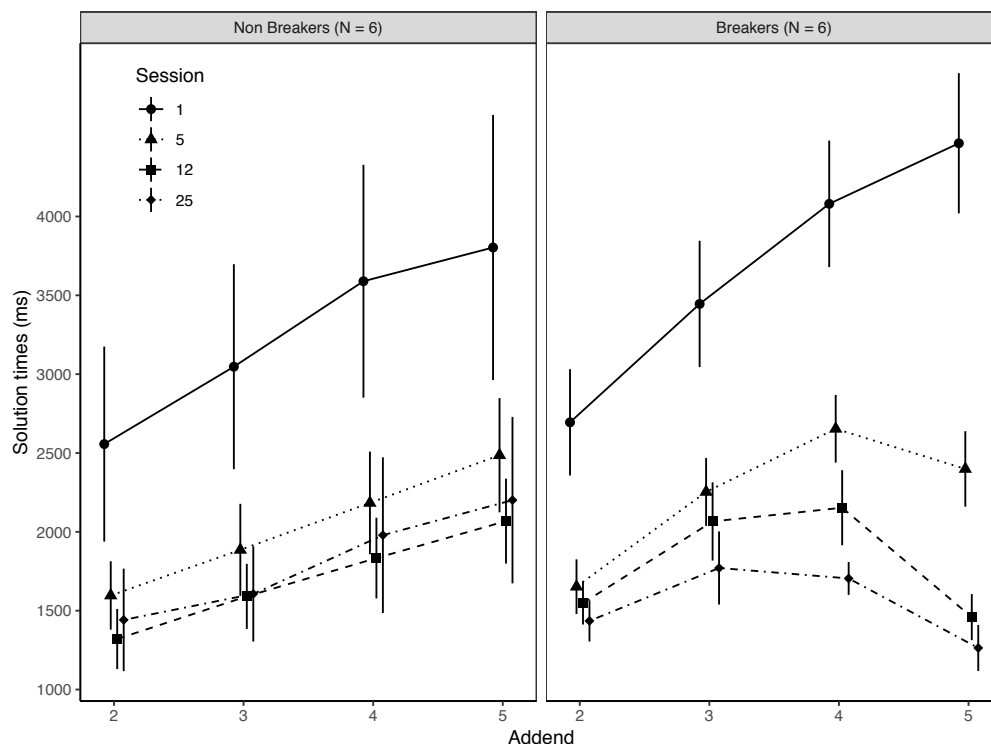
Note. Difference in solution times for true equations between problems with addends 4 and 5 across the 25 sessions, when all problems are included (circles, solid line) and when the end-term problems are excluded (triangles, dotted line). Because of large variations and for the sake of clarity, error bars are not represented in the Figure.

Interestingly, scrutinizing the whole data set at an individual level revealed that, across training sessions, not all participants showed a break in solution times at addend 4. Therefore, based on participants' solution times for +4 and +5 problems, we created two groups. Six participants who never showed a break across sessions were classified as non

breakers. Six participants who showed a systematic break from one session (from as early as in Session 1 to as late as in Session 17) until the end of training were classified as breakers. Solution times for these two groups are presented in Figure 5. The 7 remaining participants did not show such consistent pattern across sessions, because they presented aberrant behaviors with appearance and disappearance of the break across sessions. In order to ensure that our categorization into breakers and non breakers was relevant, we conducted a series of one-tailed paired-sample t -tests on solution times at Session 25 and showed that, for breakers, +4 problems were solved faster than +5 problems, ($t(5) = 5.38, p = .001$) and that, for non breakers, +4 problems were solved slower than +5 problems ($t(5) = -3.83, p = .006$).

Figure 5

Mean Solution Times as a Function of Addend



Note. Mean solution times as a function of addends for Sessions 1 (circles, solid line), 5 (triangles, dotted line), 12 (squares, dashed line), and 25 (diamonds, dot-dashed line) for true

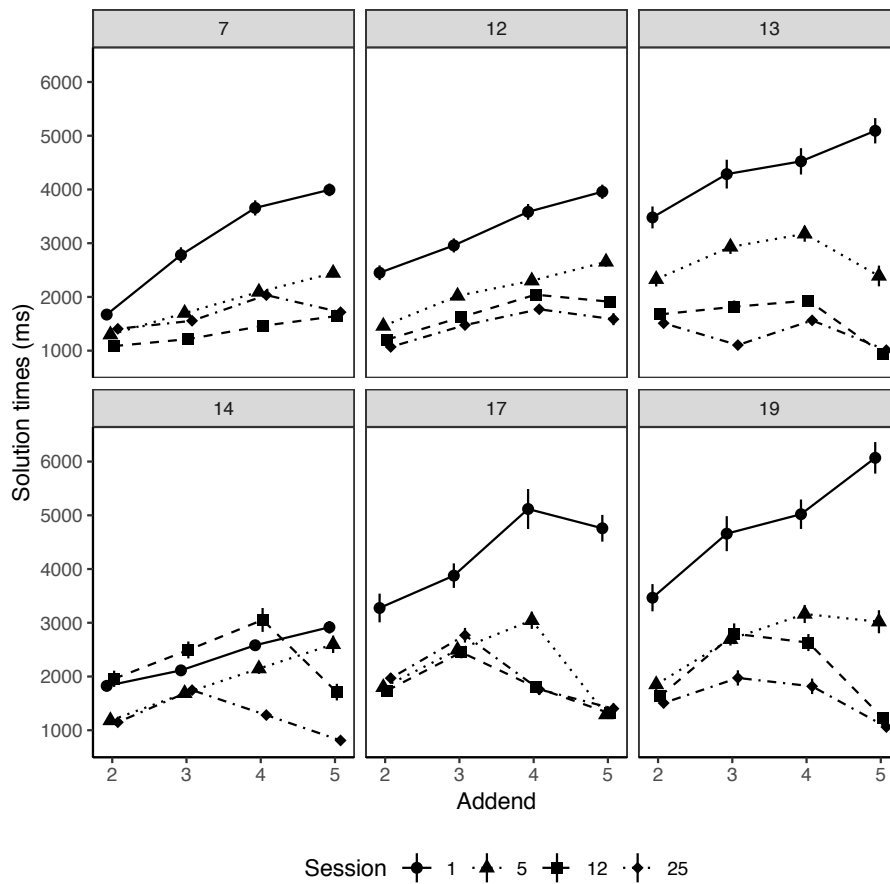
trials in non breakers (left panel) and breakers (right panel). Error bars represent standard errors.

We conducted a 4 (Addend: 2 to 5) x 2 (Group: breakers vs. non breakers) mixed-design ANOVA on solution times in Session 25, with Group as a between measure. There was no effect of Group ($F(1, 10) < 1$) but the interaction between Group and Addend was significant ($F(3, 30) = 7.24, \eta_p^2 = .42, p < .001$). A series of contrasts with Holm correction revealed that for non breakers, the effect of Addend was linear and significant ($t(10) = 3.44, p = .02$) while for the break group, the effect was not linear ($t(10) = -0.75, p = .94$) but quadratic ($t(10) = -5.54, p < .001$).

At a more descriptive level and in regards to Figure 5, it seems that +5 problems are not the only problems that are processed differently by the populations of breakers and non breakers. Whereas, in non breakers, solution times for +4 problems are consistent with the linear trend, in breakers, those problems do not conform to the linear function because they are solved quicker (1705 ms) than +3 problems (1771 ms). In order to shed light on this matter, we examined the data in breakers at an individual level (Figure 6). Interestingly, except for one participant with a peculiar pattern of solution times (i.e., participant 13), two different patterns of behavior emerged from the data. Some of the participants showed the already described discontinuity at +4 whereas some others presented the discontinuity at +3.

Figure 6

Mean Solution Times as a Function of Addend



Note. Mean solution times as a function of addends for Sessions 1 (circles, solid line), 5 (triangles, dotted line), 12 (squares, dashed line), and 25 (diamonds, dot-dashed line) for participants classified as breakers. Number above each panel indicates participant number. Error bars represent standard errors.

Discussion

This first experiment was conducted in order to address some issues about the conclusions drawn by Logan and Klapp (1991) that individuals retrieve the results of alphabet-arithmetic problems after a training experiment. To this aim, we trained 19 participants using the same material as the authors (i.e., same letters associated with same

addends from 2 to 5) over 25 instead of 12 sessions in the original study. After 12 sessions, we obtained the same pattern of results as Logan and Klapp, with an increase in response times from addends 2 to 4 and a decrease between addends 4 and 5. The same pattern was also observed after 25 sessions, which ensures that these results were not due to a lack of training. This point is important considering that, from Session 1 onwards, our 19 participants were slower than the 8 participants involved in Logan and Klapp's study. The slope in solution times depending on the addends in Session 12 was also steeper in our study (163 ms) than in the original one (45 ms). It could have been suggested that after 12 sessions, our participants were not trained enough and, consequently, that their behavior could not be compared to Logan and Klapp's results. Nevertheless, a lack of practice cannot explain our results because, as attested by the absence of statistical difference between Sessions 12 and 25, further practice does not seem to impact the solution time pattern observed at mid-training.

More central to our study, we questioned the conclusions in relation to the decrease in solution times observed for +5 problems in Logan and Klapp's (1991) study. We suggested that this phenomenon could be related to end-term effects, with endpoints in a set of stimuli being particularly salient and therefore processed faster. Our results show that removing end-term problems (i.e., J + 5 and T + 5 for A to J and K to T set of letters, respectively) attenuated the discontinuity in solution times but did not make it disappear. As evident in Figure 4, this discontinuity appears from Session 6 onwards and remains present until the end of training, even when end-term problems were excluded. Therefore, end-term effects cannot fully explain the pattern of results observed in alphabet-arithmetic tasks and our results support the interpretation of Logan and Klapp (1991) that the discontinuity in solution times for +5 problems can be due to deliberate mnemonic encoding strategies for +5 problems.

Therefore, +5 problems were not processed as the other problems during the course of the training. This is the reason why we recalculated the slopes by excluding those problems. Our analysis revealed steep positive slopes of 217 ms and a clear linear pattern of solution times as a function of addends (Figure 3). This pattern of results is compatible with the idea that participants still scan the letters of the alphabet one by one at the end of training for problems involving addends from 2 to 4. Indeed, the mental recitation rate of 115 ms per letter measured by Logan and Klapp (1991) at the end of their experiment is considerably lower than the addend slope of 217 ms/addend observed in our experiment. This point will be taken up in our General Discussion.

In addition to classical analyses at the sample level, we investigated potential differences of behavior between our participants. This approach revealed the existence of two contrasted populations. While one third of our participants presented aberrant behaviors with appearance and disappearance of the break across sessions, one third of our participants presented a break at +4 at one point of the training until the end whereas another third never presented this break. Within the breakers group, we went even deeper in individual observations and identified a sub-population of 3 participants who presented a break at +3. Highly interestingly and as observable in Figure 6, those participants presented the +3 break after they presented the +4 break during the course of training. If, as we concluded following Logan and Klapp (1991), a break in solution times is the sign of deliberate memorization, our results show that some of our participants memorize the material from the largest problems to the smallest. As far as non breakers were concerned, the slope relating solution times to addends 2 to 5 was of 265 ms/addend. Therefore, one third of participants, even after having been exposed to twice the number of problem presentations than in Logan and Klapp's experiment, present a behavior that is perfectly compatible with the idea that they still count at the end of training and never deliberately memorize any category of problems. Moreover,

at an intra-individual level, we do not exclude the possibility that breakers still count for some + 5 problems. This would explain why solution times for + 5 problems are still higher than for + 2 and + 3 problems (see Figure 2 and 3). In fact, such mixed strategies could also be evoked for problems that are solved primarily by counting strategies. We do not exclude the possibility that non breakers retrieve the results of a minority of problems. These considerations echo Siegler's overlapping waves model (1996) in which at a given point in development, several strategies can be used to solve the same problem.

All in all, our results do not show evidence against the fact that counting is still the dominant strategy at the end of an extensive alphabet-arithmetic training. When retrieval is probable, it concerns only some participants and is used preferentially for problems associated with the largest addends. A simple test of this interpretation is to present participants with problems associated with larger addends than in previous training programs. Therefore, in a second experiment, we asked participants to solve alphabet-arithmetic problem with addends from 2 to 6 instead of 2 to 5. If our understanding of occasional retrieval during alphabet-arithmetic training is correct, some participants should present a break in solution times at +5 problems rather than at +4 as observed in Experiment 1. Alternatively, it is possible that it is from +5 problems that participants decide not to count and therefore deliberately memorize +5 and +6 problems simultaneously during the course of the experiment. In this case, a decrease in solution times should be observable for +5 problems and solution times should be similar for +6 and +5 problems. Because in this first experiment, the break in solution times at addend 4 was observable from the 6th session, in the second experiment we stuck to the original number of 12 sessions used by Logan and Klapp (1991).

EXPERIMENT 2

Method

Participants. Twenty-four students (13 females) aged between 18 and 29 years were recruited by means of the student-job websites of the University of Lausanne and the Swiss Federal School of Technology in Lausanne. All participants were native French speakers and they received CHF 120 for their participation. Informed consent was obtained for each of the participant.

Material and procedure

Participants were trained on equations with addends from 2 to 6 paired with either the first 8 letters of the alphabet (A to H) or the next 8 letters (I to P). The same number of problems (i.e., 40), number of trials per session (i.e., 480) as well as number of problem presentations per session (i.e., 6 times with TRUE and 6 times with FALSE answers) as in Logan and Klapp (1991) and as in our Experiment 1 were kept. The same design as in Experiment 1 was used except that we took an additional methodological precaution and counterbalanced the response keys across participants. Half of the participants were required to press the “A” key when the presented equation was correct and the “L” key when it was incorrect while the other half were required to press the opposite keys. Moreover, and in order to better control the randomization, the 480 trials per session in the current experiment were divided into 3 identical blocks of 160 trials instead of 4 blocks with a different random mix of problems in Experiment 1. The presentation of problems was still fully randomized within each block.

In contrast to Experiment 1 and in order to reduce potential experimental noise, participants were trained at the laboratory for 12 sessions on 12 consecutive working days. In order to maximize the efficiency of training, they were required to train at home during the week end by verifying 160 equations presented on paper as quickly as possible. This corresponded to one block of the experiment.

Results

Three participants who presented flat addend slopes from the beginning of training (i.e., Session 1 or 2) were excluded from our analyses. The results of these participants could not be informative concerning our research question, which is related to the evolution of counting strategy after extensive training. These participants obviously did not count at the beginning of the study and were therefore out of the scope of our purpose.

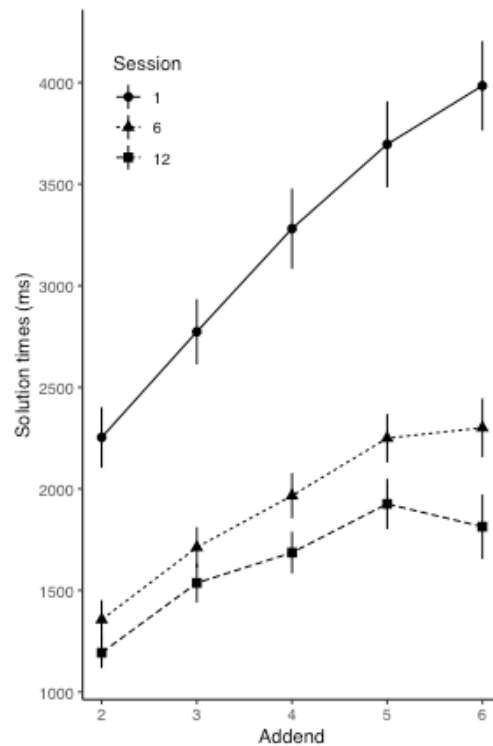
Accuracy. Overall, the percentage of correct responses was of 96% ($SD = 5\%$), from 95% in Session 1 to 96% in Session 12. A 2 (Letter Set: A to H vs. I to P) x 12 (Session: 1 to 12) x 5 (Addend: 2 to 6) mixed-design ANOVA with Letter Set as between measure was conducted on correct responses. The main effect of Letter Set was not significant ($F(1, 19) = 1.01, p = .33$) and this variable did not interact with Session or Addend ($F_s < 1$). Therefore, we collapsed the two letter sets and conducted a 12 (Session: 1 to 12) x 5 (Addend: 2 to 6) repeated-measures ANOVA. It confirmed that the rates of correct responses increased across sessions ($F(11, 220) = 2.06, \eta_p^2 = .09, p = .02$) and revealed that they decreased as a function of addends ($F(4, 80) = 9.04, \eta_p^2 = .31, p < .001$). The interaction between Session and Addend did not reach the .05 significance level ($F(44, 880) = 1.33, p = .08$). As in the first experiment, due to high percentage of correct responses, the remainder of the result section will focus on solution times.

Solution times. As in Experiment 1, we conducted our analyses on true equations only and we removed the values shorter than 300 ms and longer than mean + 3 SD per participant and per session, which resulted in 1.3% of discarded data.

A 2 (Letter Set: A to H vs. I to P) x 12 (Sessions: 1 to 12) x 5 (Addend: 2 to 6) mixed-design ANOVA with Letter Set as between measure was conducted on solution times. We did not find an effect of Letter Set or an interaction between this variable and Session or Addend (all $F_s < 1$). After collapsing the two letter sets, we conducted a 12 (Session: 1 to 12) x 5 (Addend: 2 to 6) repeated-measures ANOVA. The mean solution time across sessions was

of 2023 ms ($SD = 1056$ ms) and significantly decreased over sessions, from 3198 ms in Session 1 to 1631 ms in Session 12, $F(11, 220) = 93.93$, $\eta_p^2 = .82$, $p < .001$. A main effect of Addend was also found ($F(4, 80) = 121.91$, $\eta_p^2 = .86$, $p < .001$), showing that solution times increased as a function of addends (i.e., 1436, 1802, 2085, and 2367, and 2428 ms for Addend 2, 3, 4, 5, and 6 respectively).

More interestingly, the interaction between Session and Addend was significant ($F(44, 880) = 20.40$, $\eta_p^2 = .50$, $p < .001$, see Figure 7). Still, when the addend slope was calculated for each participant and each session, it remained significant from the beginning until the end of the training (Sessions 1: $M = 441$ ms/addend, $SE = 30$, $p < .001$; Session 6: $M = 243$ ms/addend, $SE = 20$, $p < .001$; Session 12: $M = 163$ ms/addend, $SE = 24$, $p < .001$). The difference in addend slope between Session 1 and Session 12 was significant ($t(20) = 7.99$, $p < .001$).

Figure 7*Mean Solution Times as a Function of Addend*

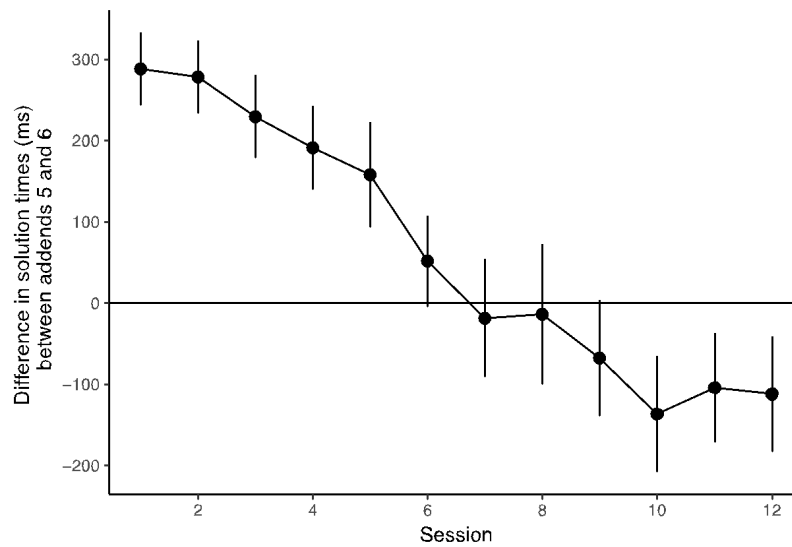
Note. Mean solution times as a function of addends for Sessions 1 (circles, solid line), 6 (triangles, dotted line), and 12 (squares, dashed line) for true trials. Error bars represent standard errors.

In order to determine whether solution times increased regularly with addends or in other words whether the effect of addend on solution times followed a linear trend, we carried out a series of contrast with Holm correction. In Session 1, the linear ($t(20) = 14.62, p < .001$) and quadratic ($t(20) = -3.80, p = .003$) trends were significant. In Session 12, both trends were still significant ($t(20) = 6.80, p < .001$ for linear trend and $t(20) = -5.27, p < .001$ for quadratic trend).

Highly interestingly and in contrast with Experiment 1, we found a discontinuity in solution times at +5 instead of at +4 problems. Stated differently, whereas in Experiment 1, solution times decreased from addend 4 to addend 5 problems, solution times in this second

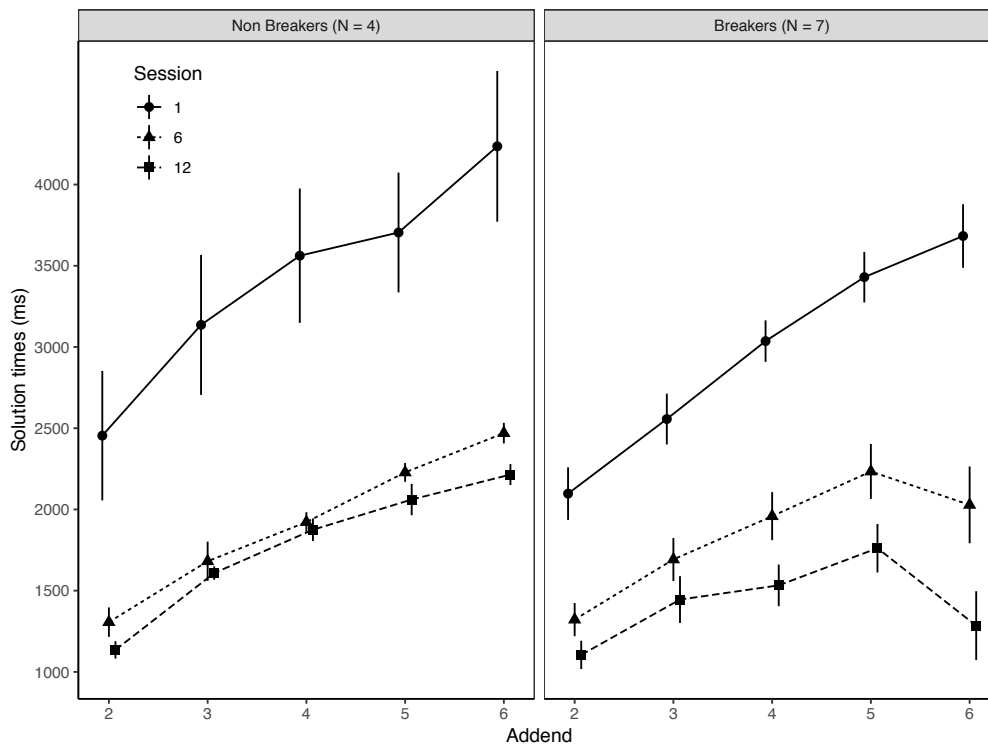
experiment decreased from addend 5 to addend 6 problems. Figure 8 shows that descriptively, solution times increased noticeably from addends 5 to 6 in the first session (+281 ms) but were closer to each other in the middle session (+48 ms) and decreased in the last session (–112 ms).

When the addend slopes were recalculated without the highest addend, we found that they were significantly steeper than when all the addends were considered. In Session 1, the slope for addends 2 to 5 corresponded to 487 ms/addend and differed significantly from the slope of 441 ms/addend for addends 2 to 6 ($t(20) = 3.53, p = .002$). In Session 6, the slope for addends 2 to 5 corresponded to 294 ms/addend and differed significantly from the slope of 243 ms/addend for addends 2 to 6 ($t(20) = 4.44, p < .001$). In Session 12, the slope for addends 2 to 5 corresponded to 236 ms/addend and differed significantly from the slope of 163 ms/addend for addends 2 to 6 ($t(20) = 5.03, p < .001$). When +6 problems were not considered, the difference in addend slope between Session 1 and Session 12 was significant ($t(20) = 6.18, p < .001$).

Figure 8*Difference in Solution Times as a Function of Session*

Note. Difference in solution times between problems with addends 5 and 6 across the 12 sessions for true equations. Error bars represent standard errors.

Similar to Experiment 1, a closer look at the data at individual level showed that not all participants showed a discontinuity in solution times. Therefore, based on participants' solution times for +5 and +6 problems, we created two groups. Four participants who never showed a break were classified as non breakers. Seven participants who showed a systematic break from one session (from as early as in Session 1 to as late as in Session 12) until the end were classified as breakers (Figure 9). The 10 remaining participants did not show such consistent pattern across sessions. In order to ensure that our categorization into breakers and non breakers was relevant, we conducted a series of one-tailed paired-sample *t*-tests on solution times at Session 12 and showed that, for breakers, +5 problems were solved faster than +6 problems, ($t(6) = 5.96, p < .001$) and that, for non breakers, +5 problems were solved slower than +6 problems ($t(3) = -1.92, p = .08$).

Figure 9*Mean Solution Times as a Function of Addend*

Note. Mean solution times as a function of addends for Sessions 1 (circles, solid line), 6 (triangles, dotted line), and 12 (squares, dashed line) for true trials for non breakers (left panel) and breakers (right panel) Error bars represent standard errors.

Discussion

This second experiment was carried out in order to better understand the fact that, at the end of an alphabet-arithmetic task training involving addends from 2 to 5, solution times associated with +5 problems are shorter than for +4 problems. The results of our first experiment confirmed Logan and Klapp's (1991) interpretation that this decrease in solution times is due, at least in part, to deliberate mnemonic encoding of +5 problems. Logan and Klapp's point of view was that retrieval for +2 to +4 problems would be possible after repetitive practice consisting in one-unit steps counting. In contrast for the same authors, some participants could consider that counting one by one for +5 problems is too demanding

and too long and would decide to directly commit their answers to long-term memory shortly after the beginning of the experiment. In order to examine this interpretation, we asked the participants of this second experiment to solve problems with addends 2 to 6 and showed that the decrease in solution times was shifted from +5 to +6 problems. Therefore, if we stick to the interpretation that a break in solution times is the sign of deliberate memorization, we can conclude that, as in the first experiment, some of the participants in this second experiment have deliberately used mnemonic strategy to learn the alphabet-arithmetic facts associated with the highest addend in the study set.

GENERAL DISCUSSION

Automaticity in the instance theory of automatization consists in a shift from algorithmic procedures to retrieval of instances from long-term memory (Logan, 1988). Learning would consist in a transition from time-consuming processing of each of the algorithmic steps to rapid single-step access to the instances. In an alphabet-arithmetic task, this would appear in a shift from a steep linear function relating solution times to the addends at the beginning of learning to a flat function at the end of learning. In this paper, we have questioned this conclusion because we argue that, in previous studies, the computation of the slope of the function at the end of learning was biased by a drop in solution times for the largest addend.

We have shown here that the systematic drop of solution times for the largest addend was reduced but still present when end-term problems were excluded from the analyses. Then, following Logan and Klapp (1991), we concluded that the residual drop in solution times could be due to deliberate memorization of problems associated with the largest addend. Therefore, the problems that would be memorized first by individuals would be those requiring the highest number of counting steps. Our first experiment conducted over 25

sessions instead of 12 showed that amongst the 6 out of 19 participants who memorized the problems with the largest addend, half of them also started to memorize problems associated with the second-largest operand. In the second experiment, conducted over 12 sessions using addends from 2 to 6 instead of 2 to 5, we showed that the decrease in solution times after training was, again, observed for the largest addend. This pattern of results was caused by 7 participants out of 21.

Because problems associated with the highest addend were obviously processed differently from other problems, we recalculated the slopes associating solution times with addends excluding the highest one. Both in Experiments 1 and 2, these slopes were not at all negligible at the end of training (217 and 235 ms/addend, respectively) and their sizes were higher than the subvocal or overt alphabet recitation time per letter (125 ms/letter in Landauer, 1962 and 115 ms/letter after training in Logan & Klapp, 1991). Therefore and contrary to Logan and Klapp's previous conclusions, it is impossible from these results to exclude the possibility that individuals are still counting for most of the problems after extensive practice. This leaves room to the possibility that training in an alphabet-arithmetic task simply increases the speed of execution of the counting steps, except for a very specific category of problems that are processed differently by a minority of participants.

The fact that repeated practice of an algorithmic procedure finally results in an increased counting speed echoes the conclusions formulated in the more-ecological context of mental arithmetic. For example, Svenson (1985) argued that simple arithmetic problems with a sum inferior to 10 and involving either 1 or 2 as addends could be solved by a counting procedure whereas larger problems could be solved by retrieval. Interestingly, this conclusion that very-small problems are solved using counting procedures whereas the answers to larger problems can be retrieved has also been reached by Uittenhove et al. (2016). These authors revealed that adults present a linear and monotonic increase in solution times for very-small

non-tie additions involving operands from 1 to 4, whereas there was no increase in solution times for problems with a sum from 7 to 10 when one of the operands exceeded 4. The conclusion of the authors was that very-small problems could be solved by very-quick and unconscious automated one-unit step procedures, as long as the operands were not too high. When operands are too high or perceived as such by individuals, the counting procedure could appear as too demanding, long and possibly prone to errors and retrieval would be favored. These conclusions perfectly fit the conclusions formulated in the present paper. In an alphabet-arithmetic task, practice would lead to an acceleration of procedures for problems requiring a minimum number of steps, whereas the answers to problems with numerous steps could be strategically committed to memory. This interpretation echoes the conclusion of Compton and Logan (1991) that the effects observed in alphabet-arithmetic tasks fit well with a race model (Logan & Cowan, 1984) in which memory traces race against the algorithms. When the algorithm is slow, as for large problems, the memory traces are more likely to win than when the algorithm is fast, as for small problems.

Svenson and Uittenhove et al. found that each automated step takes approximately 45 ms, which is far from the average slope of 250 ms that we found after 12 sessions in the present experiments. Therefore, it is very unlikely that the counting procedures used by our participants are already automated. Our hypothesis is that automatization would occur after a much longer training period than what could reasonably be done in a laboratory. Indeed, a slope of 45 ms is observable after many years of practice. As a matter of fact, even at the age of 10 years, after about 4 years of practice, children do not seem to have already automatized counting procedures (390ms/increment when the minimum addend is considered, Thevenot et al., 2016).

Of course, another possibility in the alphabet-arithmetic task would be that after many years of practice, the answers to all problems are retrieved from long-term memory.

Nevertheless, in this case, the evolution pattern of retrieval that we observed is in conflict with the current models of retrieval in arithmetic. Indeed, within these models, the answers to small problems are assumed to be retrieved before the answers to large problems during the course of development (e.g., Ashcraft, 1982; 1992; Campbell, 1995; Campbell & Oliphant, 1992; Chen & Campbell, 2018; Siegler, 1996). Several arguments support this assumption. First, associations between operands and results are easier when the number of steps to reach the correct answer is small than when it is large (Thevenot et al., 2001). Second, the associations between operands and answers are stronger when problems are less error prone and therefore stronger for small than larger problems (Siegler & Shrager, 1984). Third, the probability of associations between operands and answers is higher for problems that are encountered more often and earlier during development, that is higher for small than large problems (Ashcraft & Christy, 1995; Hamman & Ashcraft, 1986). Therefore, because we show here that the answers to large problems, and not to small problems, are retrieved first during practice, even if we speculate that counting procedures will eventually disappear, most of the retrieval models need to be revised.

To conclude, the results of the present paper challenge the conclusion that the pattern of solution times after extensive alphabet-arithmetic training rules out the possibility that fast counting cannot be the dominant strategy. In other words, we have dethroned one of the main arguments that procedures are taken over by memory retrieval after repeated practice by showing that the slopes relating solution times to addends are still not negligible after 12 or even 25 sessions of training. Of course, this does not question the whole instance theory put forward by Logan (1988), according to which extensive practice eventually allows the retrieval of associations between operands and answers. First, we have seen that, in accordance with the original conclusions drawn by Logan and Klapp, the answers to some

problems associated with the highest operands can be deliberately committed to memory by a minority of participants.

Second, in addition to the pattern of solution times after alphabet-arithmetic training, several other arguments were put forward by Logan and his colleagues. Zbrodoff (1999) revealed the existence of what she called the opportunistic-stopping phenomenon. This phenomenon supposes that when an incorrect equation is presented and that the proposed answer corresponds to a letter preceding the correct answer, individuals would not count up to the correct answer but only up to the proposed one. Consequently, verification times of such equations should be shorter than when the proposed answer follows the right answer. Accordingly, the author observed at the beginning of an alphabet-arithmetic task that True-2 problems (e.g., $A + 4 = C$) were rejected faster than True-1 problems (e.g., $A + 4 = D$) whereas at the end of training, this opportunity stopping phenomenon was no longer observed. Zbrodoff naturally concluded that retrieval instead of counting became the dominant strategy. Nevertheless, this phenomenon will need to be investigated further because in her experiment, and in sharp contrast with what could be predicted from the opportunity stopping phenomenon, True-1 problems were not rejected faster than True problems. In fact, only True-2 problems were processed quicker than True problems. Obviously, these inconsistent results relative to the opportunistic-stopping phenomenon will need to be studied further. This is precisely what we aim at doing in a future research wherein opportunistic stopping effects will be examined as a function of the addends involved in the problems. Specifically, these effects need to be scrutinized in light of the results of the present paper by taking into account the influence of the largest addend.

Next, in order to strengthen our position that fast counting procedures are still used by adults after extensive practice, our future research will need to address the question of transfer of practice. Logan and Klapp (1991) showed poor transfer of knowledge to an untrained part

of the alphabet and therefore concluded that learning is item specific. Stated differently, according to the authors, a general procedure should have been transferred to new problems and the lack of transfer constituted evidence that counting was no longer used at the end of training. However, we suggest that possible transfer of practice should be examined in the future without considering the problems that are responsible for the break in solution times because they probably do not rely on counting procedures any longer.

Finally, in Compton and Logan (1991), participants verbally reported that they increasingly relied on retrieval over the course of training. Moreover, participants reported more retrieval strategies for problems with larger addends than smaller addends. However, the evolution of strategies during the course of learning was not considered for each of the addend. This is unfortunate because, even if verbal reports are not always trustworthy (e.g., Kirk & Ashcraft, 2001; Thevenot et al., 2010), deliberate mnemonic strategies should be associated with an abrupt shift from reported counting to reported retrieval. Therefore, asking participants how they think they solve the problem in each session for each addend could be valuable in future experiments.

References

- Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, *128*, 186-197. <https://doi.org/10.1037/0096-3445.128.2.186>
- Ashcraft, M. H. (1982). The development of mental arithmetic: A chronometric approach. *Developmental Review*, *2*, 213-236. [https://doi.org/10.1016/0273-2297\(82\)90012-0](https://doi.org/10.1016/0273-2297(82)90012-0)
- Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, *44*, 75-106. [https://doi.org/10.1016/0010-0277\(92\)90051-I](https://doi.org/10.1016/0010-0277(92)90051-I)
- 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*, *5*, 396-421. <https://doi.org/10.2307/749430>
- Baroody, A. J. (1983). The development of procedural knowledge: An alternative explanation for chronometric trends of mental arithmetic. *Developmental Review*, *3*, 225-230. [https://doi.org/10.1016/0273-2297\(83\)90031-X](https://doi.org/10.1016/0273-2297(83)90031-X)
- Baroody, A. J. (1984). A reexamination of mental arithmetic models and data: A reply to Ashcraft. *Developmental Review*, *4*, 148-156. [https://doi.org/10.1016/0273-2297\(84\)90004-2](https://doi.org/10.1016/0273-2297(84)90004-2)
- Baroody, A. J. (1994). An evaluation of evidence supporting fact-retrieval models. *Learning and Individual Differences*, *6*, 1-36. [https://doi.org/10.1016/1041-6080\(94\)90013-2](https://doi.org/10.1016/1041-6080(94)90013-2)
- Baroody, A. J. (2018). A commentary on Chen and Campbell (2017): Is there a clear case for addition fact recall? *Psychonomic Bulletin & Review*, *25*, 2398-2405. <https://doi.org/10.3758/s13423-018-1440-y>
- Barrouillet, P., & Thevenot, C. (2013). On the problem size effect in small addition: 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 single addition and multiplication: A modified network-interference theory and simulation. *Mathematical Cognition, 1*, 121-164.
- Campbell, J. I. D., Chen, Y., Allen, K., & Beech, L. (2016). Transfer of training in alphabet arithmetic. *Memory & Cognition, 44*, 1288. <https://doi.org/10.3758/s13421-016-0631-x>
- Campbell, J. I. D., & Oliphant, M. (1992). Representation and retrieval of arithmetic facts: A network-interference model and simulation. In J. I. D. Campbell (Ed.), *The nature and origin of mathematical skills* (pp. 331-364). Amsterdam: Elsevier Science.
- 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>
- Compton, B. J., & Logan, G. D. (1991). The transition from algorithm to retrieval in memory-based theories of automaticity. *Memory & Cognition, 19*, 151-158. <https://doi.org/10.3758/BF03197111>
- 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>
- Foos, P. W., Smith, K. H., Sabol, M. A., & Mynatt, B. T. (1976). Constructive processes in simple linear-order problems. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 759-766.
- Geary, D. C. (1996). The problem-size effect in mental addition: developmental and cross-national trends. *Mathematical Cognition, 2*, 63-93. <https://doi.org/10.1080/135467996387543>
- Groen, G. J., & Parkman, J. M. (1972). A chronometric analysis of simple addition. *Psychological Review, 79*, 329-343. <https://doi.org/10.1037/h0032950>

Hamann, S., & Ashcraft, M. H. (1986). Textbook presentations of the basic addition facts.

Cognition and Instruction, 3, 173-202. https://doi.org/10.1207/s1532690xci0303_2

Kirk, E. P., & Ashcraft, M. K. (2001). Telling stories: The perils and promise of using verbal

reports to study math strategies. *Journal of Experimental Psychology: Learning,*

Memory and Cognition, 27, 157-175. <https://doi.org/10.1037/0278-7393.27.1.157>

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

<https://doi.org/10.2466/pms.1962.15.3.646>

Li, M., Liu, D., Li, M., Dong, W., Huang, Y., & Chen, Q. (2018). Addition and subtraction

but not multiplication and division cause shifts of spatial attention. *Frontiers in Human*

Neuroscience, 12, 183. <https://doi.org/10.3389/fnhum.2018.00183>

Liu, D., Cai, D., Verguts, T., & Chen, Q. (2017). The time course of spatial attention shifts in

elementary arithmetic. *Scientific Reports*, 7, 921. [https://doi.org/10.1038/s41598-017-](https://doi.org/10.1038/s41598-017-01037-3)

[01037-3](https://doi.org/10.1038/s41598-017-01037-3)

Logan, G.D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95,

492-527. <https://doi.org/10.1037/0033-295X.95.4.492>

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-](https://doi.org/10.1037/0033-295X.91.3.295)

[295X.91.3.295](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, 179-195. [https://doi.org/10.1037/0278-](https://doi.org/10.1037/0278-7393.17.2.179)

[7393.17.2.179](https://doi.org/10.1037/0278-7393.17.2.179)

Masson, N., Letesson, C., & Pesenti, M. (2018). Time course of overt attentional shifts in

mental arithmetic: Evidence from gaze metrics. *Quarterly Journal of Experimental*

Psychology, 71, 1009-1019. <https://doi.org/10.1080/17470218.2017.1318931>

- Mathieu, R., Gourjon, A., Couderc, A., Thevenot, C., & Prado, J. (2016). Running the number line: Rapid shifts of attention in single-digit arithmetic. *Cognition, 146*, 229-239.
<https://doi.org/10.1016/j.cognition.2015.10.002>
- Mathieu, R., Epinat-Duclos, J., Léone, J., Fayol, M., Thevenot, C., & Prado, J. (2018). 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., Thevenot, C., & Prado, J. (2018). What's behind a "+" sign? Perceiving an arithmetic operator recruits brain circuits for spatial orienting. *Cerebral Cortex, 28*, 1673-1684.
<https://doi.org/10.1093/cercor/bhx064>
- Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature, 215*(5109), 1519-1520. <https://doi.org/10.1038/2151519a0>
- Potts, G. R., Banks, B. P., Kosslyn, S. M., Moyer, R. S., Riley, C. A., & Smith, K. H. (1978). Encoding and retrieval in comparative judgments. In N. J. Castellan, Jr., & F. Restle (Eds.), *Cognitive theory* (pp. 243-308). Hillsdale, NJ: Lawrence Erlbaum.
- Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. Oxford: Oxford University Press.
- Siegler, R. S., & Jenkins, E. (1989). *How children discover new strategies*. Hillsdale, NJ: Lawrence Erlbaum.
- Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 31-76). Hillsdale, NJ: Erlbaum.

- 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.
- 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](https://doi.org/10.1016/0001-6918(85)90048-4)
- 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., Barrouillet, P., & Fayol, M. (2001). Algorithmic solution of arithmetic problems and operands-answer associations in LTM. *Quarterly Journal of Experimental Psychology-A*, *54*, 599-611. <https://doi.org/10.1080/02724980042000291>
- Thevenot, C., Castel, C., Fanget, M., & Fayol, M. (2010). Mental subtraction in high and lower-skilled arithmetic problem solvers: Verbal report vs. operand-recognition paradigms. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *36*, 1242-1255. <https://doi.org/10.1037/a0020447>
- Thorndike, E. (1911). *Animal intelligence: Experimental studies*. New York: Macmillan.
- 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>
- Zbrodoff, N. J. (1999). Effects of counting in alphabet arithmetic: Opportunistic stopping and priming of intermediate steps. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 299-317. <https://doi.org/10.1037/0278-7393.25.2.299>

Zhou, X. L., Chen, C. S., Zang, Y. F., Dong, Q., Chen, C. H., Qiao, S. B., & Gong, Q. Y.

(2007). Dissociated brain organization for single-digit addition and multiplication.

Neuroimage, 35, 871-880. <https://doi.org/10.1016/j.neuroimage.2006.12.017>

Zhu, R., Luo, Y., You, X., & Wang, Z. (2018). Spatial bias induced by simple addition and

subtraction: From eye movement evidence. *Perception*, 47, 143-157.

<https://doi.org/10.1177/0301006617738718>

Conflict of interest

We declare no conflict of interest

Ethical principles

We declare that this research has complied with APA ethical standards in the treatment of our human sample. We have obtained approval from the Research Committee of the Social and Political Sciences Faculty of University of Lausanne as well as approval from the local committee of the Psychology Department at the University of Geneva. According to the Swiss Federal Act on Research involving Human Beings and because our study is only behavioral and conducted on a non-pathological population, we are exempted from submitting the research proposal to the Cantonal Ethic Committee.