Journal of Experimental Child Psychology 265 (2026) 106457

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

Patterns of solution times for simple addition problems in 12-year-old children are not compatible with retrieval models: A rebuttal to Andras and Macizo (2025)

Catherine Thevenot [a,*], Jérôme Prado [b,**]

[a] *Institut de Psychologie, Université de Lausanne, Switzerland*
[b] *Centre de Recherche en Neurosciences de Lyon (CRNL), INSERM U1028 - CNRS UMR5292, Université de Lyon, France*

A R T I C L E I N F O

Keywords:
Arithmetic
Retrieval models
Automatized counting theory
Mental strategies

A B S T R A C T

The cognitive mechanisms supporting arithmetic learning, in simple addition in particular, have long been debated. Whereas traditional models propose that counting strategies are gradually replaced by direct retrieval over development, the automatized counting theory suggests that expert performance on very small additions rather relies on rapid, unconscious one-by-one counting procedures. A central point of disagreement between these accounts concerns the interpretation of the problem-size effect, namely the increase in solution times for larger addition problems. Retrieval models attribute this effect to increased interference among arithmetic facts in memory, whereas the automatized counting theory posits that the effect reflects the larger number of counting steps required for larger problems. Nevertheless, recent findings showing that the size effect disappears for sums beyond 7 challenge the interference account, which would predict a monotonic increase in solution times with increasing problem size. Rather, these findings are consistent with the automatized counting theory, according to which problems involving operands greater than 4 fall outside the range of automatization. However, Andras and Macizo (2025) recently reported a failure to replicate this breakpoint at sum 7 in 6th graders and concluded that the monotonic increase of solution times they observe support retrieval-based accounts. Nonetheless, an examination of their data indicates that this conclusion is not empirically supported. In reality, Andras and Macizo (2025)'s results reveal a non-monotonic pattern, with a lack of problem size effect for problems with sums beyond 7, which is inconsistent with retrieval-based interference models and aligns with the automatized counting theory.

There has been a long-standing debate regarding the strategies used by expert solvers when faced with single-digit addition problems. Several decades ago, most researchers agreed that the counting procedures used by children in early stages of learning are gradually replaced by the direct retrieval of answers from long-term memory in adolescents and adults (Ashcraft, 1982; 1992; Ashcraft & Stazyk, 1981; Siegler & Shrager, 1984). However, it is also possible that arithmetic fluency in experts might instead, or at least also, rely on the automatization of numerical schemas, rules, and principles (Baroody, 1983; 1984; 1994). Baroody's idea that procedures may become progressively automatized rather than supplanted by memory retrieval has been recently revitalized by the automatized counting theory, proposed by Thevenot and her collaborators (e.g., Fayol & Thevenot, 2012; Barrouillet & Thevenot, 2013; Thevenot

* Address: Institut de Psychologie, Université de Lausanne, Géopolis, CH-1015 Lausanne, Switzerland.
** Address: Centre de Recherche en Neurosciences de Lyon, CH Le Vinatier, 95 bd Pinel, 69675 Bron Cedex, France.
E-mail addresses: catherine.thevenot@unil.ch (C. Thevenot), jerome.prado@cnrs.fr (J. Prado).

https://doi.org/10.1016/j.jecp.2026.106457
Received 14 December 2025; Received in revised form 5 January 2026;
Available online 16 January 2026

et al., 2016).

According to the automatized counting theory, experts solve small addition problems involving operands up to 4 by relying on very fast, one-by-one counting processes that occur automatically and outside conscious control (Uittenhove et al., 2016). This procedure would be triggered and executed to completion as soon as the problem is presented and even cued by the mere appearance of an arithmetic symbol such as a '+' sign (e.g., Mathieu, Epinat-Duclos, Sigovan, et al., 2018). There is now substantial evidence for the automatized counting theory. Evidence comes from both brain imaging (e.g., Díaz-Barriga Yáñez et al., 2023; Mathieu, Epinat-Duclos, Léone, et al., 2018) and behavioral studies in adults and children, using cross-sectional and longitudinal designs across various experimental paradigms (e.g., see Chouteau et al., 2024 or Thevenot et al., 2020 for the use of an alphabet-arithmetic task; Díaz-Barriga Yáñez et al., 2020 or Mathieu et al., 2016 for the spatial manipulations of numbers; or Poletti et al., 2021 for a priming paradigm). However, because the automatized counting theory contradicts long-standing associationist views of arithmetic learning, it has quite naturally sparked intense debate within the scientific community (e.g., Baroody, 2018; Campbell et al., 2021; Chen & Campbell, 2017; Prado & Thevenot, 2021; Thevenot & Barrouillet, 2020).

One of the most contentious issues in this debate concerns how the problem-size effect observed with small addition problems (i.e., problems with sums up to 10) is interpreted in expert solvers. The problem-size effect refers to the finding that solution times increase with problem size. For example, 2 + 3 is typically solved faster than 4 + 3, which is also solved faster than 5 + 3 (e.g., LeFevre et al., 1996). The problem-size effect is only observed for non-tie problems (i.e., problems with two different operands) and is among the most robust findings in the numerical cognition literature (Zbrodoff & Logan, 2005). The effect is easy to interpret when children still rely on conscious and slow counting strategies, since larger problems involve more counting steps and therefore longer solution times (Groen & Parkman, 1972). However, adults also show a problems size effect despite not using explicit counting procedures for these small problems with sums up to 10. According to the automated counting theory, which assumes that counting may be accelerated and become unconscious with expertise, the problem size effect would still be due to counting. In other words, it may still stem from an increase in counting steps in expert solvers (these steps would be so fast that they escape consciousness). Nonetheless, the problem-size effect is more difficult to reconcile with retrieval models, and proponents of this theory must invoke peculiarities of interconnected memory networks of arithmetic facts to account for it (Campbell, 1987). More specifically, within retrieval models, the problem-size effect is often thought to arise from interference among arithmetic facts within this type of network. As the number combinations increase in size, they tend to share more overlapping features or associations with other facts (e.g., similar operands or results), which would lead to greater competition during retrieval. This interference would slow down access to the correct answer, resulting in longer response times for larger than smaller problems (Campbell & Graham, 1985).

Because the problem-size effect can be explained by both the automated counting and retrieval theories, researchers face a challenge in using the effect to discriminate between the two accounts. Nevertheless, the discovery by Uittenhove et al. (2016) that the size effect was specific to problems with sums up to 7 and disappeared for sums of 8, 9, and 10 reignited the debate (see Fig. 1A), placing the problem-size effect once again at the center of discussions among researchers (Baroody, 2018; Chen & Campbell, 2021; Thevenot & Barrouillet, 2020). This plateau in solution times from sums of 8 to sums of 10, replicated several times across many experiments in different population of adults and children from the age of 7 (Bagnoud et al., 2025; 2021; Díaz-Barriga Yáñez et al., 2023; Poletti et al., 2023; Thevenot et al., 2016), is a major challenge for retrieval models. Indeed, if the size effect for small problems is

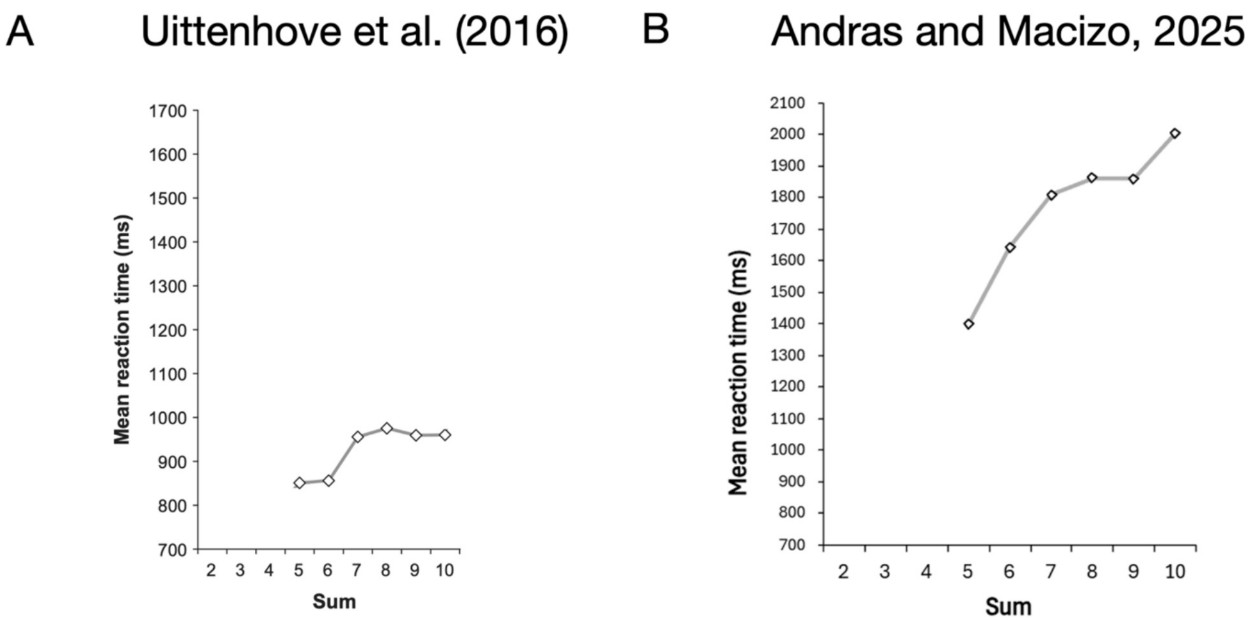

Fig. 1. Means RTs as a function of problem sum for non-tie problems from sum of 5 to sum of 10 in adults (from Uittenhove et al., 2016) (A) and 6th-graders (from Andras and Macizo, 2025) (B).

due to longer retrieval times due to interference, it should be expected that solution times increase monotonically across problems that are supposedly solved through retrieval. Any break in this progression necessarily challenges the interpretation of the size effect as reflecting greater difficulty in retrieving larger rather than smaller problems (because of increased interferences). On the contrary, this plateau provides additional support for the automated counting theory. This is because automated counting may only operate within the subitizing range, i.e., when quantities can be represented within a single focus of attention. There is extensive evidence that the upper limit of this range is four items (Cowan, 2001; Kaufman et al., 1949). Therefore, automatized counting may only operate when operands are smaller than (or equal to) 4. This predicts an increase in response time as problem size increases up to the sum of 7 (e.g., 4 + 3), due to an increase in counting steps. However, beyond that range (i.e., for problem with a sum of 8, such as 5 + 3) automatized counting may no longer operate, which would explain the plateau observed for problems with sums from 8 to 10 in previous studies.

However, in a recent study, Andras and Macizo (2025) raised doubts regarding the existence of a break point in addition solution times at the sum of 7, followed by a plateau from sum of 8 to sum of 10. As noted earlier, this pattern had previously been replicated across many experiments, including a longitudinal study tracking 51 children from Grade 3 to Grade 5 (Bagnoud et al., 2021). However, Andras and Macizo claimed that they do not find statistical evidence for the existence of a plateau from sums 8 to 10 in a population of 28 sixth graders. They therefore concluded that there is a monotonic increase of solution times for non-tie problems with sums from 5 to 10, which contradicts the automatized counting model and supports retrieval-based accounts. Andras and Macizo's data are available in the Open Science Framework repository at https://osf.io/38kdz/overview.

There are nonetheless several reasons to question the conclusion from Andras and Macizo (2025). First, from an empirical perspective, a simple inspection of Andras and Macizo's data in 6th-graders, reproduced here on Fig. 1B (thanks to the open dataset from the authors) only for non-tie problems from sum of 5 to sum of 10 (the original plot can be seen on Andras and Macizo's own Fig. 1), reveals that Andras and Macizo's claim may not be correct. Solution times in 6th-graders for sum of 3 to sum of 10 do not appear to follow a monotonic pattern. In fact, as is made clear by comparing Fig. 1B to Fig. 1A, the pattern Andras and Macizo obtained in 6th-graders is not unlike the pattern obtained by Uittenhove et al. (2016) with adults. That is, the increase in solution times from sum of 5 to sum of 10 is not monotonic, either in Uittenhove et al. (2016) or Andras and Macizo (2025). Rather, in both dataset, response times appear to increase from sum of 5 to sum of 7, after which a plateau is reached at least until sum of 9.

Second, to quantify the patterns they obtain, Andras and Macizo rightly split their problems with a sum up to 10 between those in which operands are smaller than (or equal to) 4 (which they term 'very-small problems') and problems in which operands are larger than 4 (which they term 'medium-small problems'). This is appropriate for testing the automatized counting model because, according to that model, only very-small problems are candidate for procedural automatization and therefore should show a problem size effect in 6th graders. As is clear from Fig. 2 (generated thanks to the open dataset from the authors) both categories of problems show a clear

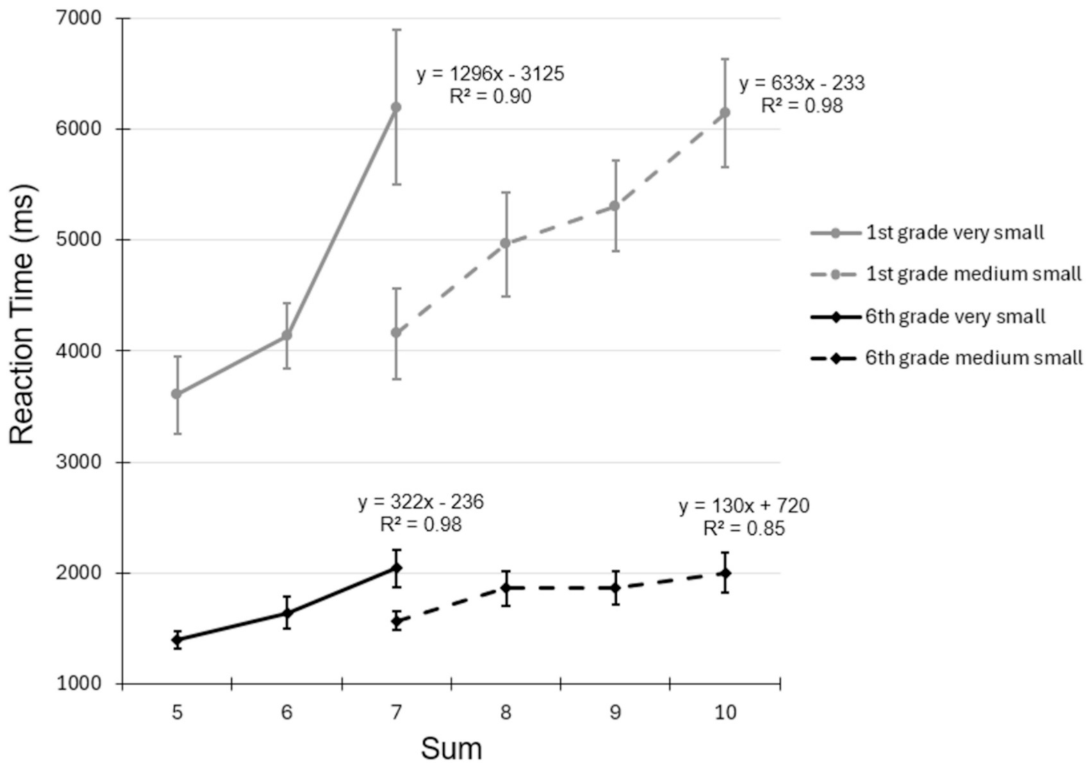

Fig. 2. Means RTs as a function of problem sum for non-tie problems, split by category (very-small and medium-small) for 1st-graders and 6th-graders (from Andras and Macizo, 2025).

problems size effect in 1st graders, which is expected given that young children use explicit counting procedures. However, solution times for medium small problems do not increase to the same extent as for small problems in 6th graders. In fact, Andras and Macizo themselves report in 6th-graders (i) a larger problem-size effect for very-small problems (slope = 0.13 log(ms)) than for medium-small problems (slope = 0.05 log(ms)) and (ii) a non-significant increase in solution times for medium-small problems ($t(29.07) = 1.88$, $p = .07$) whereas the effect is highly significant for very small problems ($t(28.26) = 4.89$, $p < .001$) (Andras and Macizo, 2025, p. 9 and p. 10). In other words, Andras and Macizo's own statistics are not consistent with their conclusion that "non-tie problems always presented a problem-size effect [...] in the sixth-grade children" (Andras and Macizo, 2025, p. 1).

To provide deeper insight into the pattern actually measured in 6th-graders, we extracted from Andras and Macizo's open dataset the response times for each problem category and sum in 6th-graders. In very-small problems, paired t-tests revealed that problems with a sum of 5 were faster than problems with a sum of 6 ($t(27) = -2.09$, $p = .046$), which were in turn faster than sum of 7 problems ($t(27) = -2.84$, $p = .009$). In other words, a clear problem-size effect is observed in this problem category. In medium-small problems, problems with a sum of 7 were faster than problems with a sum of 8 ($t(27) = -2.89$, $p = .007$). However, no difference was observed between problems with a sum of 8 and problems with a sum of 9 ($t(27) = .02$, $p = .982$) and between problems with a sum of 9 and problems with a sum of 10 ($t(27) = -1.61$, $p = .119$). In other words, there was a plateau from problems with a sum of 8 to problems with a sum of 10 in medium-small problems, exactly as reported in Uittenhove et al. (2016) and contrary to Andras and Macizo's interpretation. It is also interesting to observe that, in their dataset, problems with a sum of 7 are solved faster when they belong to the medium-small than to the very-small category ($t(27) = 3.46$, $p = .002$) (see Fig. 2). This difference is simply impossible to explain with a model that places interferences at the source of the problem size effect, as both types of problems should have the same level of interference. Overall, this difference and the much weaker (or absent) size effects for the medium-small than the very-small category make the conclusion drawn by Andras and Macizo not supported by their data.

To conclude, and as already discussed in several articles reporting the existence of a plateau from sums 7 (Bagnoud et al., 2025, 2021; Díaz-Barriga Yáñez et al., 2023; Poletti et al., 2023; Thevenot et al., 2016; Uittenhove et al., 2016), Andras and Macizo (2025)'s results do appear to be consistent with the automatized counting theory. As noted above, automatized counting is assumed to operate only for problems in which both operands are smaller than (or equal to) 4. These problems are therefore particularly susceptible to a pronounced size effect, reflecting one-by-one counting steps. In contrast, other small problems beyond this range, previously referred to as medium problems and referred to as medium-small problems, could be immune to size effects because they cannot be solved through automatized counting and are instead likely solved via retrieval from long-term memory.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Tatum Holley for assisting with data analysis and visualization.

Data availability

Andras and Macizo's data are available in the Open Science Framework repository at https://osf.io/38kdz/overview.

References

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

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

Ashcraft, M. H., & Stazyk, E. H. (1981). Mental addition: A test of three verification models. *Memory & Cognition, 9*, 185–196. https://doi.org/10.3758/BF03202334

Bagnoud, J., Dewi, J., Castel, C., Mathieu, R., & Thevenot, C. (2021). 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*, Article 104987. https://doi.org/10.1016/j.jecp.2020.104987

Bagnoud, J., Poletti, C., Krenger, M., Mahendrathas, M., Dewi, J., & Thevenot, C. (2025). Unraveling the tie problem mystery in addition: From finger counting to mental strategies in first graders. *Journal of Experimental Child Psychology, 252*, Article 106154. https://doi.org/10.1016/j.jecp.2024.106154

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. 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

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 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. (1987). The role of associative interference in learning and retrieving arithmetic facts. In J. A. Sloboda, & D. Rogers (Eds.), *Cognitive Processes in Mathematics* (pp. 107–122). Clarendon Press/Oxford University Press.

Campbell, J. I. D., & Chen, Y. (2021). Not toeing the number line for simple arithmetic: Two large-n conceptual replications of Mathieu et al. (Cognition, 2016, Experiment 1). *Journal of Numerical Cognition, 7*, 248–258. https://doi.org/10.5964/jnc.6051

Campbell, J. I. D., & Graham, D. J. (1985). Mental multiplication skill: Structure, process, and acquisition. *Canadian Journal of Psychology, 39*, 338–366. https://doi.org/10.1037/h0080065

Chen, Y., & Campbell, J. I. D. (2017). "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

Chouteau, S., Lemaire, B., Thevenot, C., Dewi, J., & Mazens, K. (2024). Learning basic arithmetic: A comparison between rote and procedural learning based on an artificial sequence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 50*, 418–434. https://doi.org/10.1037/xlm0001241

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–185. https://doi.org/10.1017/s0140525x01003922.

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 during single-digit arithmetic. *Annals of the New York Academy of Sciences, 1477*, 79–90. https://doi.org/10.1111/nyas.14464

Díaz-Barriga Yáñez, A., Longo, L., Chesnokova, C., Poletti, C., Thevenot, C., & Prado, J. (2023). Neural evidence for procedural automatization during cognitive development: Intraparietal response to changes in very-small addition problem-size increases with age. *Developmental Cognitive Neuroscience, 64*, Article 101310. https://doi.org/10.1016/j.dcn.2023.101310

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *The American Journal of Psychology, 62*, 498–525. https://doi.org/10.2307/1418556

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

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

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

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, 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: Rapid shifts of attention in single-digit arithmetic. *Cognition, 146*, 229–239. https://doi.org/10.1016/j.cognition.2015.10.002

Poletti, C., Díaz-Barriga Yáñez, A., Prado, J., & Thevenot, C. (2023). The development of simple addition problem solving in children: Reliance on automatized counting or memory retrieval depends on both expertise and problem size. *Journal of Experimental Child Psychology, 234*, Article 105710. https://doi.org/10.1016/j.jecp.2023.105710.

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

Prado, J., & Thevenot, C. A. C. (2021). A Commentary on: Not toeing the number line for simple arithmetic: Two large-n conceptual replications of Mathieu et al. (Cognition, 2016, Experiment 1). *Journal of Numerical Cognition, 7*, 473–475. https://doi.org/10.5964/jnc.7391

Siegler, R. S., & Shrager, J. (1984). A model of strategy choice. In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229–293). Hillsdale, NJ: Erlbaum.

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

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., & 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.