



Cognitive Science 45 (2021) e13074

© 2021 The Authors. *Cognitive Science* published by Wiley Periodicals LLC on behalf of Cognitive Science Society (CSS).

ISSN: 1551-6709 online

DOI: 10.1111/cogs.13074

Automatization through Practice: The Opportunistic-Stopping Phenomenon Called into Question

Jasinta D. M. Dewi, Jeanne Bagnoud, Catherine Thevenot

Institute of Psychology, University of Lausanne

Received 23 May 2021; received in revised form 9 November 2021; accepted 16 November 2021

Abstract

As a theory of skill acquisition, the instance theory of automatization posits that, after a period of training, algorithm-based performance is replaced by retrieval-based performance. This theory has been tested using alphabet-arithmetic verification tasks (e.g., is $A + 4 = E?$), in which the equations are necessarily solved by counting at the beginning of practice but can be solved by memory retrieval after practice. A way to infer individuals' strategies in this task was supposedly provided by the opportunistic-stopping phenomenon, according to which, if individuals use counting, they can take the opportunity to stop counting when a false equation associated with a letter preceding the true answer has to be verified (e.g., $A + 4 = D$). In this case, such within-count equations would be rejected faster than false equations associated with letters following the true answers (e.g., $A + 4 = F$, i.e., outside-of-count equations). Conversely, the absence of opportunistic stopping would be the sign of retrieval. However, through a training experiment involving 19 adults, we show that opportunistic stopping is not a phenomenon that can be observed in the context of an alphabet-arithmetic verification task. Moreover, we provide an explanation of how and why it was wrongly inferred in the past. These results and conclusions have important implications for learning theories because they demonstrate that a shift

A part of the results has been presented at the European Society for Cognitive Psychology Conference, Tenerife (Spain), 25–28 September 2019.

This research was supported by the Swiss National Foundation for Scientific Research through the grant 100014_138185 to Catherine Thevenot.

Correspondence should be sent to Catherine Thevenot, Institute of Psychology, University of Lausanne, Géopolis, Lausanne 1015, Switzerland. E-mail: catherine.thevenot@unil.ch

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

from counting to retrieval over training cannot be deduced from verification time differences between outside and within-count equations in an alphabet-arithmetic task.

Keywords: Learning; Knowledge acquisition; Training; Strategies; Counting; Retrieval

1. Introduction

Many activities, such as reading, writing, driving, recalling autobiographical memories, or solving simple additions, can be performed by a majority of adults without considerable cognitive effort. This leads many researchers to the conclusion that individuals perform them by memory retrieval. More generally, it is widely accepted in the literature that the end product of a learning process often consists of retrieval of associations, irrespective of the way the knowledge is acquired. Some knowledge, such as the name of capitals, can be learnt directly and deliberately by the memorization of associations (e.g., Norway–Oslo or Ecuador–Quito), while other knowledge, such as associations between operands and answers in addition problems (e.g., $4 + 3 = 7$ or $7 + 6 = 13$), can be created after repeated practice of counting procedures (e.g., $4 + 3 = 5, 6, 7$). In the latter case, a shift from procedural to retrieval strategies necessarily occurs during learning (e.g., Ashcraft, 1982, 1992; Campbell, 1995; Campbell & Oliphant, 1992; Chen & Campbell, 2018; Siegler, 1996).

In the domain of arithmetic, and especially in mental additions, the proponents of retrieval models infer the shift from procedures to retrieval from the evolution of solution times in the course of development and practice (e.g., Ashcraft & Battaglia, 1978; Logan & Klapp, 1991). In fact, irrespective of the developmental stage, solution times for simple addition problems increase with the size of the smaller operand involved in the problem (e.g., Zbrodoff & Logan, 2005), and this was believed to be due, among other factors, to smaller problems being practiced more often than larger ones. However, the slope of the regression line decreases drastically with age, from 400 ms/increment in 6-year-old (Groen & Parkman, 1972) to 260 ms/increment in third graders and 120 ms/increment in sixth graders (Jerman, 1970), and finally 20 ms/increment in adults (Parkman & Groen, 1971). The slope of 400 ms/increment in 6-year-olds is viewed as reflecting their counting speed because it is unlikely that they already rely on retrieval for nontie problems. However, the reduction of the size of the slopes in older individuals can be the product of both accelerated counting speed and increased use of memory retrieval. Nevertheless, Groen and Parkman naturally considered the possibility that the slope of 20 ms/increment found in adults corresponds only to counting speed but considered it to be too fast, particularly in comparison to the overt or subvocal recitation speed, that is, 125 ms/letter (Landauer, 1962). Therefore, they concluded that adults generally retrieve the answer of simple additions from memory but fail to do so in about 5% of the problems. In more recent models, such size effects are interpreted by better memory access and therefore, shorter retrieval times for smaller problems that are learnt earlier (Ashcraft, 1982) and more frequently (e.g., Ashcraft & Christy, 1995) during development, and would hence suffer from less interference than larger ones (e.g., Campbell, 1987; 1995; Campbell & Graham, 1985).

In all retrieval models of mental addition, the shift from counting to memory retrieval during development has been explained by the strengthening of associations between operands and answers through repeated practice of counting procedures (e.g., Geary, 1996; Siegler & Jenkins, 1989; Siegler & Shipley, 1995; Siegler & Shrager, 1984). This shift from procedural- to retrieval-based performance in the acquisition of simple addition has been modeled by the instance theory of automatization (Logan, 1988). The theory purports that, at the beginning of learning, due to the lack of memory trace associated with the newly encountered material, the task is accomplished by the use of algorithm-based procedures. Then, with each instance of learning, one memory trace associated with this instance is created. The probability that a task will be performed by memory retrieval increases with the number of traces in memory. Therefore, according to the instance theory of automatization, at one point during the learning process, the probability of using memory retrieval will exceed the probability of using algorithms. According to the author, this point corresponds to the so-called automatization. From this point onward, memory retrieval will be used predominantly and involuntarily.

The instance theory of automatization was established to explain the acquisition of several cognitive skills, including the acquisition of addition through the alphabet-arithmetic task (Compton & Logan, 1991; Logan & Klapp, 1991), where a combination of a letter augend and a numerical addend results in a letter answer, for example, $A + 5 = F$. Using this task, Logan and Klapp trained adult participants to verify alphabet-arithmetic equations involving addends 2, 3, 4, and 5 over 12 days. On the first training day, the slope of solution times as a function of addend (hereinafter: addend slope) was 486 ms/addend, indicating the use of a counting strategy. In the last training session, the addend slope decreased to 45 ms/addend. The reduction in addend slope found by Logan and Klapp can be paralleled to the abovementioned results of Groen and Parkman (1972), who obtained a slope of regression line for the smaller operand of 400 ms/increment in children and 20 ms/increment in adults. The small and nonsignificant addend slope of 45 ms/addend at the end of training was one of Logan and Klapp's arguments that a shift from counting to memory retrieval had taken place.

However, this argument has recently been questioned by Thevenot, Dewi, Bagnoud, Uitenhove, and Castel (2020). In fact, irrespective of the addends used in the study, a systematic phenomenon has been observed in alphabet-arithmetic studies, namely that at the end of training, there is a discontinuity in the increase of solution times as a function of addend (e.g., Chen, Orr, & Campbell, 2020; Compton & Logan, 1991; Dewi, Bagnoud, & Thevenot, 2021; Logan & Klapp, 1991; Wenger, 1999; Zbrodoff, 1995, 1999). More precisely, in Logan and Klapp's work involving addends from 2 to 5, solution times increase from addends 2 to 4 and then decrease for addend 5. This systematic finding has led Thevenot et al. to argue that the decrease of addend slope is mainly due to the problems with the largest addend, that is, solution times problems involving 5 were lower than for problems involving 4. They demonstrated this by excluding problems with the largest addend from the analyses and obtaining a significant addend slope at the end of training. However, this effect involving problems with the largest addend was only observed in a minority of participants (6 out of 19 in their Experiment 1 and 7 out of 21 in their Experiment 2). Therefore, Thevenot et al. concluded that the nonsignificant addend slope at the end of training in Logan and colleagues' experiments

must have resulted from the averaging of solution times across participants, which artificially reduced the addend slope. Furthermore, they logically concluded that memorization of associations between operands and answers in an alphabet-arithmetic task occurs only for the largest problems and for a minority of participants, even after extensive practice. Obviously, these conclusions stand in opposition to Logan's (1988) instance theory of automatization. However, Logan's theory is not based solely on the size of the addend slope at the end of learning.

1.1. The opportunistic-stopping phenomenon

Another signature of a shift from counting to retrieval in alphabet-arithmetic verification tasks was proposed by Zbrodoff (1999). She analyzed different false equations of the same problem and argued that if counting were used to verify a false equation associated with an answer preceding the true answer (i.e., within-count answer, e.g., $A + 5 = E$), then participants would stop counting once the proposed answer has been reached. In other words, in verifying the false equation $A + 5 = E$, the participant would stop counting at E and would not continue all the way to the correct answer F. In contrast, if a false equation associated with an answer following the true answer (i.e., outside-of-count answer, e.g., $A + 5 = G$) is presented, then participants would count up to either the correct answer or the proposed answer. As a consequence, within-count equations would be rejected faster than outside-of-count equations. Zbrodoff called this phenomenon opportunistic stopping and the present paper aims at investigating it further.

In Experiment 1 conducted over one session, Zbrodoff (1999) asked her participants to verify 54 problems, consisting of 18 letters paired with addends 3, 4, and 5. Each problem appeared four times with their true answer (T) and four times with false answers. When a false answer was presented, it could correspond to two letters following the true answer (T+2), the letter following it (T+1), the letter preceding it (T-1), or two letters preceding it (T-2). The author concluded that opportunistic stopping was observed, because, on average, within-count equations were rejected 560 ms faster than outside-of-count equations. Furthermore, T-2 equations were rejected 434 ms faster than T-1 equations. The author's results are reproduced here on the left panel of Fig. 1.

In her Experiment 4, Zbrodoff (1999) used the same paradigm as in her Experiment 1 but with a fewer number of problems, more repetitions of the same problem, and over 13 training sessions instead of a single one. More precisely, participants had to verify 27 problems consisting of nine letters paired with addends 3, 4, and 5. In each session, each problem appeared eight times with their true answer and eight times with false answers. The author observed that within-count equations were rejected 401 ms faster than outside-of-count equations on the first day. Because opportunistic stopping was obtained, the author concluded that counting strategy must have been used. Furthermore, although Zbrodoff did not report the values, it is clear from the right panel of Fig. 1 (that replicates Zbrodoff's fig. 3) that T-2 equations were rejected faster than T-1 equations, replicating the results of her first experiment. However, opportunistic stopping disappeared on day 13 (see the right panel of Fig. 1). Stated differently, rejection times for within-count and outside-of-count equations were similar. This

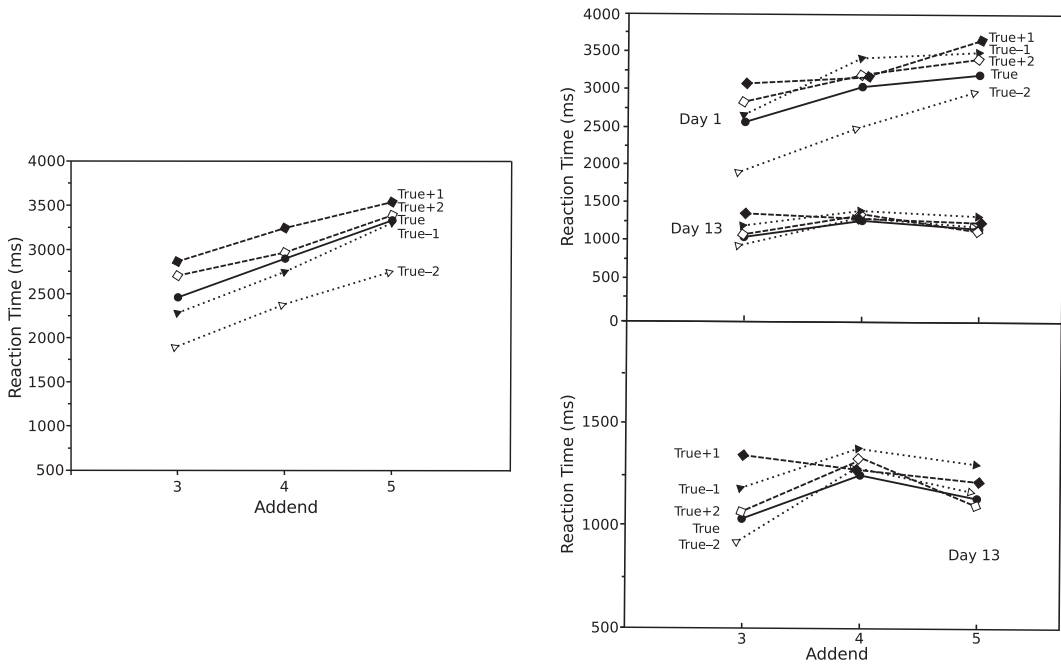


Fig 1. Solution times as a function of addend.

Note. Solution times as a function of addends in Experiment 1 (left panel) and Experiment 4 (right panel) of Zbrodoff (1999). Results for day 13 of Experiment 4 are enlarged on the bottom panel. Solid circles and solid lines represent solution times for T equations, open lozenges and dashed lines for T+2 equations, solid lozenges and dashed lines for T+1 equations, solid triangles and dotted lines for T-1 equations, and open triangles and dotted lines for T-2 equations. Adapted from “Effects of Counting in Alphabet Arithmetic: Opportunistic Stopping and Priming of Intermediate Steps”, by N. J. Zbrodoff, 1999, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, p. 303 (left panel) and p. 311 (right panel). Copyright 1999 by the American Psychological Association.

disappearance was interpreted as the evidence that a shift from counting to memory retrieval had taken place.

Thus, using opportunistic stopping, Zbrodoff (1999) adduced additional evidence for the shift from counting to retrieval in tasks initially requiring procedural algorithms. Nevertheless, detailed inspections of her results, particularly of her Experiment 4, cast doubts on her conclusions. As we will explain in the following subsection, her results were not consistent with her definition of opportunistic stopping. Furthermore, an alternative explanation can account for her results.

1.2. Why opportunistic stopping does not correspond to the opportunity to stop

If opportunistic stopping reflects the use of counting strategy, then not only should solution times for T-2 equations be shorter than for any other equations but also solution times for

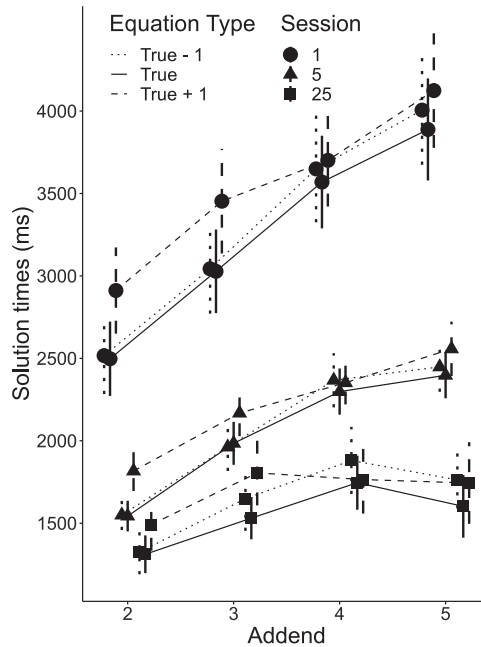


Fig 2. Solution times as a function of addend and equation type.

Note. Solution times as a function of addend and equation type (solid lines for T, dotted lines for T-1, and dashed lines for T+1 equations) in sessions 1 (solid circles), 6 (solid triangles), and 25 (solid squares). Error bars represent standard errors.

T-1 equations should be shorter than for T and outside-of-count equations (i.e., T+1 and T+2). Following this rationale, T-2 equations were indeed solved faster than other equations in Experiment 1 and on day 1 of Experiment 4. However, T-1 equations were not solved faster than T or outside-of-count equations. In fact, on day 1 of her Experiment 4, when counting was supposed to be the dominant strategy, solution times for T-1 were longer than for T equations, irrespective of the problem addend. Moreover, rejection times for T-1 were longer than for T+1 equations for +4 problems and were longer than for T+2 equations for +4 and +5 problems (see the right panel of our Fig. 1). Because these observations concerning T-1 equations are not consistent with the idea that participants stop counting once the proposed answer is reached, the validity of the opportunistic-stopping phenomenon as the signature of a counting strategy is questionable. This is highly problematic because if the existence of opportunistic stopping at the beginning of practice cannot be taken as the sign of counting, then its disappearance at the end of practice cannot be taken as the sign of retrieval.

Interestingly, Zbrodoff (1999) found that the disappearance of opportunistic stopping occurred at the same moment as when the addend slope for T equations reached the asymptotic, nonsignificant value of 60 ms/addend (i.e., session 5). This correspondence supports the view that the disappearance of opportunistic stopping is indeed the sign of a shift from

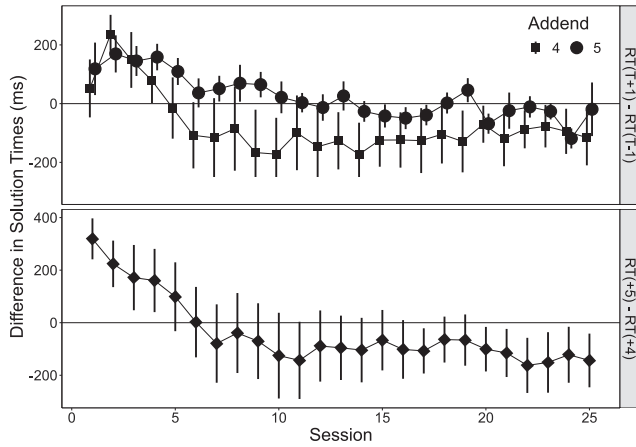


Fig 3. Difference in solution times across sessions.

Note. Top panel: Difference in solution times between T+1 and T-1 equations for +4 (squares) and +5 (circles) problems across sessions. Positive differences imply larger solution times for T+1 than for T-1. Bottom panel: Difference in solution times between +5 and +4 problems for T equations across sessions. Positive differences imply larger solution times for +5 than for +4 problems. Error bars represent standard errors.

counting to retrieval. However, as discussed earlier, the size of this nonsignificant slope was artificially lowered by problems with the largest addend +5, which were solved 110 ms faster than +4 problems. As it can be seen on the right panel of our Fig. 1, lower solution times for +5 than for +4 problems were obtained for the five types of equations (i.e., T, T-1, T-2, T+1, and T+2). Thus, it is possible that, similar to the nonsignificant addend slopes at the end of training, the disappearance of opportunistic stopping was also caused by problems with the largest addend. This interpretation is supported in Zbrodoff’s study by the fact that T+2 equations for +5 problems were solved faster than any other equations (see the bottom part of the right panel of our Fig. 1). These problems correspond to what Thevenot et al. (2020) called the end-term problems, that is, problems that have unique combination of letter augend and letter answer in the study set, that are partially responsible for the decrease in solution times. For example, for the letter augend A, the T+2 equation $A + 5 = H$ can be recognized quickly because it is the only equation pairing A and H. The salience of problems with a unique problem–answer combination may lead individuals to memorize and process them faster than nonunique problems. Note that the end-term problems also involve T-2 equations for +3 problems. For example, the equation $A + 3 = B$ is the only one pairing A and B. As will be explained later, these types of equations can also be solved quickly by the so-called letter-after strategy.

1.3. How to explain Zbrodoff (1999)’s results

After demonstrating that the so-called opportunistic stopping does not necessarily reflect the use of counting strategy in Zbrodoff’s (1999) experiment, we propose alternative

explanations for her results concerning, first, shorter rejection times for T-2 than for other problems in day 1 and, second, the disappearance of opportunistic stopping in day 13.

Zbrodoff (1999) found that rejection times for T-2 equations in day 1 were shorter than for other equations. As already extensively explained, her interpretation was that participants counted until they reached the proposed answer and then stopped. However, the use of other strategies than counting can also explain this effect. For example, participants might have used plausibility judgments, which do not require counting and can lead to short solution times (Lemaire & Fayol, 1995; Lemaire & Reder, 1999; Masse & Lemaire, 2001; Reder, 1982; Zbrodoff & Logan, 1990). It turns out that T-2 equations can easily give rise to such judgments. First, the proposed answer for T-2 problems is relatively close to the letter augend (e.g., $C + 4 = E$ or $A + 5 = D$) and, without counting, it is easy to figure out that E cannot be four letters apart from C or D cannot be five letters apart from A. Second, to solve T-2 problems, participants can use a letter-after strategy, which would be similar to the “number after N ” strategy described in mental arithmetic (Baroody, 1995; Baroody, Eiland, Purpura, & Reid, 2012). More precisely, to solve $N + 1$ or $1 + N$ problems, instead of counting one step, individuals can simply retrieve the next number after N in the counting sequence (Bagnoud et al., 2021; Bagnoud, Dewi, Castel, Mathieu, & Thevenot, 2021; Grabner, Brunner, Lorenz, Vogel, & De Smedt, 2021). This highly salient relation between numbers or letters of the alphabet can allow individuals to immediately realize that, for example, $G + 3 = H$ is false because H immediately follows G in the alphabet and cannot, therefore, be three letters apart (Table A.1). Nevertheless, such plausibility judgments can also be applied to T-1 equations (e.g., $G + 2 = H$), and we have seen that they are not processed faster than other equations. Another rule that individuals could use to avoid counting would be to use the “skip the number after N ” strategy instead of counting two steps (Baroody, 2018). For example, the T-2 equation $A + 4 = C$ can be judged quickly as incorrect because C is reachable by skipping the letter after A and hence cannot be four letters apart from A (Table A.1). More generally, such strategies, which do not imply one-by-one counting, can be applied when the proposed false answer is close to the letter used in the problems, and this could explain why within count answers were rejected faster than outside of count equations.

Therefore, the so-called opportunistic-stopping phenomenon supposed by Zbrodoff (1999) in day 1 does not necessarily reflect the use of counting. Hence, its disappearance in day 13 does not necessarily imply a shift from counting to retrieval. An explanation of this disappearance can be found in a reversal, or change of sign from positive to negative, of the differences in rejection times between within-count and outside-of-count equations for +4 and +5 problems in day 13 (see the bottom part of the right panel of Fig. 1). More precisely, for +3 problems, both within-count equations were rejected faster than the two outside-of-count equations. For +4 problems, T-1 were rejected slower than T+1 but T-2 were rejected faster than T+2. For +5 problems, however, both within-count equations were rejected slower than the two outside-of-count equations. Although the results for +3 problems and for T-2 of the +4 problems could be explained by the letter-after or skip-letter-after strategies illustrated in Table A.1, the results for +5 problems and for T-1 of +4 problems cannot. Nevertheless, we argue that averaging the differences between within-count and outside-of-count equations over addends resulted in a difference, which was close to 0 and not significant, which

was interpreted by Zbrodoff as the disappearance of opportunistic stopping. Although her ANOVAs included addend and equation type, Zbrodoff did not report the interaction between the two variables. We think that this interaction is in fact important because it could highlight the reversal of the differences in rejection times between within-count and outside-of-count equations, particularly for problems with the largest addend. Furthermore, we argue that it is this reversal for problems with the largest addend, and not the shift to retrieval, that was responsible for the disappearance of opportunistic stopping at the end of practice. The reason why T+2 were rejected faster than T-2 for +5 problems in day 13 might be related to the fact that these equations contain the only combination of letter augend and letter answer in the study set. For example, the combination of letters A and H is found only once in the study set, that is, as the T+2 equations of the letter augend A, and hence, this combination could be recognized as a false equation faster than other combinations for +5 problems. In other words, similar to solution-time discontinuity found in earlier alphabet-arithmetic studies (e.g., Thevenot et al., 2020), the reversal of rejection times was also caused by problems with the largest addend.

Thus, to sum up, we argue that the obtained opportunistic stopping in day 1 of Experiment 4 of Zbrodoff (1999) cannot be strictly associated with the use of counting but could be related to the use of other strategies based on plausibility judgments. Furthermore, we argue that the disappearance of opportunistic stopping in day 13 is not necessarily associated with the use of retrieval but could be due to the reversal of the difference in rejection times between within-count and outside-of-count equations that is observed for problems with the largest and, to a certain extent, second-largest addends.

1.4. *The present study*

Therefore, in the present paper, we will reinvestigate the disappearance of opportunistic stopping by hypothesizing that it is due to the reversal of the sign of the difference in rejection times between within-count and outside-of-count equations. Because we do not think that Zbrodoff's results were due to opportunistic stopping, we will use the term Equation Type effect to qualify the general effect of solution times, that is, the difference between true, within-count, and outside-of-count equations. The term of opportunistic stopping will only be used when we refer to the definition put forward by Zbrodoff (1999), that is, shorter rejection times for within-count than for outside-of-count equations.

We predict that the disappearance of the Equation Type effect at the end of practice is due to problems with the largest addend, more precisely due to the reversal of the sign of the difference between within-count and outside-of-count rejection times. Moreover, because this reversal and solution-time discontinuity are both related to problems with the largest addend through the so-called end-term effect (Thevenot et al., 2020), we predict that these two phenomena will start to occur around the same time. However, because the discontinuity in solution times due to problems with the largest addend was only observed in a minority of participants (Thevenot et al., 2020), we predict that the Equation Type effect will be observed only for these participants.

In order to test our hypotheses, we used the data collected by Thevenot et al. (2020) in their first experiment. The following Method section is, therefore, the same as described for Experiment 1 in Thevenot et al., which was an alphabet-arithmetic verification training experiment ran over 25 sessions. Ten letters were paired with addends 2–5 and in each session, and each problem was presented six times with its true answer and six times with the false answers. In our experiment, we only used T–1 and T+1 equations. This way, although the use of the letter-after strategy is possible, that is, for +2 problems, the possibility of using the skip-the-letter-after strategy is reduced compared with the design used by Zbrodoff because, in our design, it is possible only for +3 problems (Table A.1). We, therefore, place ourselves in a situation where reliance of the opportunistic-stopping strategy is enhanced.

Although only the results on true equations were analyzed and presented in Thevenot et al., true and false equations were analyzed for the present study. The results reported in this paper have, therefore, never been reported elsewhere.

2. Method

2.1. Participants

Nineteen students of the University of Geneva, aged between 18 and 35 years, were recruited. A compensation of CHF 200 was offered for their participation. In order to increase their motivation, participant with the best performance during the training phase was awarded with a bonus of CHF 50.

Written informed consent was obtained for each participant. All procedures performed in this study, involving human participants, have been conducted in compliance with the Swiss Law on Research involving human beings. Because only behavioral data were collected in a nonvulnerable population of adults, the approval of the local ethic committee was not required. The study was carried out in accordance with the recommendations of the Ethics Committee of the University of Geneva, following the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

2.2. Material and procedure

The experiment was constructed as a training study of an alphabet-arithmetic verification task, wherein equations consisting of a digit addend and a letter augend (e.g., $A + 2 = D$; false) have to be verified. One half of the presented equations were associated with the correct letter (e.g., $A + 3 = D$), whereas the other half of the equations were associated with a false answer (e.g., $C + 4 = H$). Half of the false answers corresponded to the letter before the correct answer (T–1 equations) and the other half corresponded to the letter after it (T+1 equations). Participants were trained on 40 problems consisting of addends 2–5 associated with 10 letters. Half of the participants were trained with the first 10 letters of the alphabet (i.e., A to J) and the other half with the next 10 letters (i.e., K to T). Participants were randomly assigned to the first or second part of the alphabet. Exactly as in Zbrodoff's experiments, participants were instructed to solve the equations as fast as they

could while keeping the accuracy high. Keeping instructions similar to previous experiments was important because insisting only on speed could lead to an increased use of retrieval, whereas insisting only on accuracy could result in an increased use of counting (Wilkins & Rawson, 2011).

The task was created using E-prime 2.0 software. Material was set up on participants' laptops for home training. Although it can be argued that home training could involve more distractions than a laboratory training and hence might result in more noise in the data, this is not necessarily the case. In fact, the results of Experiments 1 and 2 reported in Thevenot et al. (2020) were comparable, even though the training for Experiment 1 was carried out at home and the training for Experiment 2 at the laboratory.

Each trial began with a fixation point (*) presented for 500 ms, followed by the equation. The equations were presented horizontally in Courier New font of size 18 and black color. They were positioned in the center of the screen and remained on the screen until a response key was pressed. Participants were asked to press the "A" key when the presented equation was correct and the "L" key when it was not. This information was given on the screen and participants had to remember the appropriate keys throughout the experiment session. Then, the screen remained blank for 1500 ms.

Every combination of letters (A to J or K to T) and addends (2–5) was presented six times with its true answer and six times with a false answer per session. The six false answers corresponded to three T–1 and three T+1 answers. Thus, every session involved 480 trials (i.e., 10 letters \times 4 addends \times 2 possible answers \times 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. Because participants were allowed to have a 1-day break during the week, the 25 training sessions took place over 25–30 days.

3. Results

The following results were based on correct trials only, that is, trials correctly identified as true or false, which constituted 98% of the data. We discarded further 1.6% of the correct trials due to the solution times being either too short (i.e., shorter than 300 ms) or too long (i.e., longer than the mean plus three standard deviations for each participant in each session). Fig. 2 shows the solution times as a function of addend for the three types of equations for sessions 1, 5, and 25. Following Zbrodoff (1999), we will present our analyses for the first and last sessions. Additionally, we also conducted analyses for session 17. This is because this session corresponds to the session where the exact same number of T equations (i.e., 104) as at the end of Zbrodoff's (1999) Experiment 4 has been presented to our participants.

3.1. Disappearance of opportunistic stopping

A 4 (Addend: 2, 3, 4, and 5) \times 3 (Equation Type: T–1, T, and T+1) repeated measures ANOVA was carried out on solution times of correctly solved problems in session 1. The

Table 1
Difference in solution times between different equation types

Session	Addend	T-1 versus T		T versus T+1		T-1 versus T+1	
		Δt	p	Δt	p	Δt	p
1	2	19	.82	-414	.001	-395	.009
	3	15	.80	-425	.002	-410	.002
	4	80	.80	-132	.18	-52	.80
	5	117	.28	-236	.15	-119	.28
17	2	22	.68	-245	.002	-223	< .001
	3	109	.12	-221	.01	-112	.07
	4	147	.07	-21	.78	126	.48
	5	192	.01	-153	.01	39	.25
25	2	13	.79	-180	.02	-167	.006
	3	118	.25	-274	.03	-157	.002
	4	135	.13	-19	.81	116	.41
	5	161	.007	-142	.17	19	.82

Note. Difference in solution times (Δt) between T-1 and T, between T and T+1, and between T-1 and T+1 equations and the corresponding p value. Positive differences indicate that solution times for the first equations were longer than for the second. For the sake of visibility, differences with $p < .05$ are written in bold italics.

ANOVA revealed significant main effects of Addend ($F(3, 54) = 98.92, \eta^2_p = .85, p < .001$) and Equation Type ($F(2, 36) = 14.77, \eta^2_p = .45, p < .001$). After a Holm correction, a series of contrasts revealed that T equations (3246 ms) were solved faster than T+1 equations (3548 ms, $t(18) = -4.57, p < .001$) and T-1 equations (3304 ms) were solved faster than T+1 equations ($t(18) = -3.94, p = .002$). The interaction between Addend and Equation Type was significant ($F(6, 108) = 2.89, \eta^2_p = .14, p = .01$). The same pattern of results was observed for all addend (i.e., T+1 > T-1 > T) but a series of contrasts with a Holm correction showed that the differences between T and T-1 and between T-1 and T+1 did not reach significance for +4 and +5 problems (Table 1).

A 4 (Addend: 2, 3, 4, and 5) \times 3 (Equation Type: T-1, T, and T+1) repeated measures ANOVA was carried out on solution times of correctly solved problems in session 17. The ANOVA revealed significant main effects of Addend ($F(3, 54) = 15.38, \eta^2_p = .46, p < .001$) and Equation Type ($F(2, 36) = 8.02, \eta^2_p = .31, p = .001$). After a Holm correction, a series of contrasts revealed that T equations (1559 ms) were solved faster than T-1 (1676 ms, $t(18) = -2.84, p = .02$) and T+1 equations (1719 ms, $t(18) = -3.41, p = .01$). The interaction between Addend and Equation Type was significant ($F(6, 108) = 4.58, \eta^2_p = .20, p < .001$). A series of contrasts with a Holm correction revealed different patterns for different addends (Table 1). For both +4 and +5 problems, T-1 equations were solved descriptively slower than T+1 equation, resulting in positive difference in rejection times between T-1 and T+1 equations.

A 4 (Addend: 2, 3, 4, and 5) \times 3 (Equation Type: T-1, T, and T+1) repeated measures ANOVA was carried out on solution times of correctly solved problems in session 25. The ANOVA revealed significant main effects of Addend ($F(3, 54) = 9.79, \eta^2_p = .35, p < .001$)

and Equation Type ($F(2, 36) = 4.57, \eta^2_p = .20, p = .02$). However, a series of contrasts with a Holm correction failed to reveal a difference between different equation types. The interaction between Addend and Equation Type was significant ($F(6, 108) = 3.41, \eta^2_p = .16, p = .004$) and a series of contrasts with a Holm correction revealed the same patterns as in session 17 (Table 1). Again, for both +4 and +5 problems, T-1 equations were solved descriptively slower than T+1 equation, resulting in positive difference in rejection times between T-1 and T+1 equations.

We also calculated the addend slopes for each participant and each session. The average addend slopes in session 1 were 506, 472, and 389 ms/addend for T-1, T, and T+1 equations, respectively ($ps < .001$). In session 25, they decreased to 155 ms/addend ($p < .001$) for T-1, 108 ms/addend ($p = .03$) for T, and 72 ms/addend ($p = .13$) for T+1 equations. When +5 problems were excluded, the addend slopes in session 25 were 278 ms/addend ($p < .001$) for T-1, 217 ms/addend ($p < .001$) for T, and 137 ms/addend ($p = .007$) for T+1 equations.

3.2. Opportunistic stopping and end-term effects

We predicted that the reversal of the difference in rejection times between within-count and outside-of-count equations for problems with the largest addend would coincide with the discontinuity in solution times. However, Fig. 3 shows that although solution-time discontinuity occurred in session 6 (i.e., negative difference between +5 and +4 problems), the reversal of the difference in rejection times for +5 problems took place in a later session, that is, session 11. Instead, we found that the emergence of solution-time discontinuity coincided with the reversal of the difference in rejection times for +4 problems.

3.3. Breakers and nonbreakers

As already explained, Thevenot et al. (2020) classified their participants according to whether or not they showed a solution time discontinuity or, in other words, whether solution times for problems with the largest addend were shorter than for problems with the second-largest addend. Following Thevenot et al., six participants who showed a systematic solution-time discontinuity 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, whereas six participants who did not show solution times discontinuity throughout the experiment were classified as nonbreakers. To test whether the effect of Equation Type differs in breakers and nonbreakers, a 4 (Addend: 2, 3, 4, and 5) \times 3 (Equation Type: T-1, T, and T+1) \times 2 (Group: breaker vs. nonbreakers) repeated-measures, mixed-design ANOVA was carried out on solution times of correctly solved problems in sessions 1, 17, and 25. The results for sessions 1 and 25 are presented in Fig. 4 and the results for the whole experiment in Appendix B.

In session 1, the same general results as for the whole sample were found, that is, significant effects of Addend ($F(3, 30) = 60.56, \eta^2_p = .86, p < .001$) and Equation Type ($F(2, 20) = 9.85, \eta^2_p = .50, p = .001$), as well as significant interaction between Addend and Equation Type ($F(6, 60) = 2.82, \eta^2_p = .22, p = .02$). There was no main effect of Group ($F(1, 10) < 1$) or interaction between Group and Addend ($F(3, 30) = 1.09, p = .37$), Group and Equation Type ($F(2, 20) < 1$), or Group \times Addend \times Equation Type ($F(6, 60) = 1.06, p = .40$).

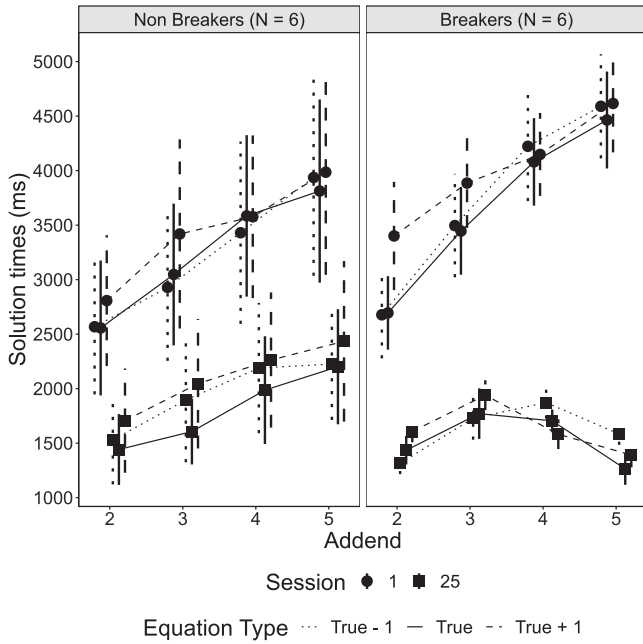


Fig 4. Solution times as a function of addend and equation type for breakers and nonbreakers.

Note. Solution times as a function of addend and equation type (solid lines for T, dotted lines for T-1, and dashed lines for T+1 equations) in sessions 1 (solid circles) and 25 (solid squares) for nonbreakers (left panel) and breakers (right panel). Error bars represent standard errors.

In session 17, the same general results as for the whole sample were found, that is, significant effects of Addend ($F(3, 30) = 9.46, \eta^2_p = .49, p < .001$) and Equation Type ($F(2, 20) = 4.02, \eta^2_p = .29, p = .03$), as well as significant interaction between Addend and Equation Type ($F(6, 60) = 5.34, \eta^2_p = .35, p < .001$). The effect of Group was not significant ($F(1, 10) < 1$), but the interaction between Group and Addend was ($F(3, 30) = 7.13, \eta^2_p = .42, p < .001$). However, a series of contrasts with a Holm correction failed to reveal a group difference in any addends. The interaction between Group and Equation Type was not significant ($F(2, 20) = 1.11, p = .35$), but the three-way interaction was ($F(6, 60) = 4.86, \eta^2_p = .33, p < .001$). A series of contrasts with a Holm correction revealed that there was an interaction between Addend and Equation Type in breakers ($F(6, 60) = 9.92, \eta^2_p = .50, p < .001$) but not in nonbreakers ($F(6, 60) < 1$). The interaction in breakers was due to T-1 rejected faster than T+1 for +2 problems ($-303 \text{ ms}, t(10) = -4.72, p = .002$) and T+1 rejected faster than T-1 for +4 problems ($+618 \text{ ms}, t(10) = -4.15, p = .005$).

In session 25, we found a significant effect of Addend ($F(3, 30) = 8.40, \eta^2_p = .46, p < .001$), a marginal effect of Equation Type ($F(2, 20) = 3.19, \eta^2_p = .24, p = .06$), and a significant interaction between Addend and Equation Type ($F(6, 60) = 2.37, \eta^2_p = .19, p = .04$). The effect of Group was not significant ($F(1, 10) < 1$), but the interaction between

Group and Addend was ($F(3, 30) = 8.66, \eta^2_p = .46, p < .001$). Again, a series of contrasts with a Holm correction failed to reveal a group difference in any addends. The interaction between Group and Equation Type was not significant ($F(2, 20) = 1.02, p = .38$), but the three-way interaction was ($F(6, 60) = 2.39, \eta^2_p = .19, p = .04$). A series of contrasts with a Holm correction revealed that there was an interaction between Addend and Equation Type in breakers ($F(6, 60) = 4.04, \eta^2_p = .29, p = .002$) but not in nonbreakers ($F(6, 60) < 1$). The interaction in breakers was due to T-1 being rejected faster than T+1 for +2 problems (-292 ms, $t(10) = -3.78, p = .009$) and +3 problems (-212 ms, $t(10) = -2.77, p = .048$), and T being solved faster than T-1 for +5 problems (-320 ms, $t(10) = -4.47, p = .003$).

4. Discussion

This research aimed at investigating some of the mechanisms at hand in cognitive learning. In the framework of the instance theory of automatization (Logan, 1988), it has been put forward that learning can correspond to a shift of strategy from algorithm-based to memory retrieval. In the alphabet-arithmetic learning, support for this theory has been provided by the opportunistic-stopping phenomenon (Zbrodoff, 1999), according to which within-count answers (e.g., $A + 4 = D$) are rejected faster than outside-of count answers (e.g., $A + 4 = F$) at the beginning of practice but not at the end. From these effects, Zbrodoff concluded that a shift from counting to retrieval occurred during the training program. Nevertheless, we noticed that Zbrodoff's results were not always consistent with her definition of opportunistic stopping, namely that even at the beginning of practice, T-1 equations were not solved faster than T equations. Indeed, in our experiment, although we confirmed the results of Zbrodoff by obtaining both an effect of Equation Type and shorter rejection times for T-1 than T+1 equations in session 1, opportunistic stopping was not observed for T compared to T-1 equations because solution times for these equations were similar. Therefore, although counting was used, participants did not take the opportunity to stop counting on reaching the within-count answer, or, at least, the paradigm used by us and Zbrodoff could not provide any evidence that opportunistic stopping was used.

At the end of practice, confirming the results of Zbrodoff, we did not find a difference in rejection times between within-count and outside-of-count equations. We could have concluded as Zbrodoff that there was no opportunistic stopping at the end of practice but in fact, it was due to a reversal of the sign of the difference in rejection times for +4 and +5 problems. More precisely, in our last session, the difference in solution times between T-1 and T+1 equations was negative for problems with addends 2 and 3 (i.e., -167 and -157 ms, respectively) but positive for problems with addends 4 and 5 (i.e., +116 and +19 ms, respectively). Averaging positive and negative differences resulted in a nonsignificant difference of -47 ms at the end of training. This reversal of the difference in rejection times was already observed in our session 17 that corresponded to Zbrodoff's day 13 in terms of the number of repetitions for each equation. It is crucial to note that the same opposite differences for problems with addend 3 on the one hand and 4 and 5 on the other hand were also obtained by Zbrodoff in her Experiment 4 (see the bottom part of the right panel of our Fig. 1) and that this is the reason

why she erroneously concluded that opportunistic stopping was absent at the end of training. In our present work, we found that the reversal for +4 problems occurred in an earlier session than for +5 problems. This might be the reason why the reversal between T+1 and T-1 in Zbrodoff's Experiment 4 was already present in day 1.

Thus, in the present paper, although we did not include T+2 and T-2 equations, we obtained the same results as Zbrodoff (1999) in her Experiment 4. First, the effect of equation type at the beginning of practice was not accompanied by shorter solution times for T-1 than for T equations. Second, within-count equations were not rejected faster than outside-of-count at the end of practice. However, we differed from Zbrodoff in the interpretation of the results. Although Zbrodoff concluded that the similarity in rejection times was due to the shift from counting to retrieval, we think, as already explained, that this is due to the reversal of the sign of the difference between the two types of false equations for +4 and +5 problems. We advance further that this reversal was mainly due to the problems with the largest addend. This proposition is supported by the fact that this reversal was only observed in breakers, for whom problems with the largest addend were processed differently than the other problems (Thevenot et al., 2020), and not in nonbreakers (see Appendix B for the performance during the whole training experiment).

To explain the reversal, we showed in Fig. 5 the performance of breakers and nonbreakers in session 25. In coherence with Table A.1, we added an example for the letter augend A. First of all, Fig. 5 shows that for breakers, T-1 equations for +2 and +3 problems were rejected faster than T+1 equations. As explained in the Introduction and illustrated in Table A.1, this can be explained by plausibility judgments. However, neither letter-after nor skip-letter-after strategy could explain why, still for breakers, T-1 equations for +4 and +5 problems were rejected slower than T+1 equations. We proposed, therefore, the following explanation. As stated by Thevenot et al. (2020), problems with the largest addend, for example, $A + 5$, are special because their letter answers appear less frequently than other letter answers in the study set. However, only the breakers are sensitive to these problem particularities and commit them to memory. Indeed, the bottom panel of Fig. 5 shows that in breakers, solution times for +5 problems were shorter than for +4 problems for the three equation types, which replicated the results of Zbrodoff (1999) in day 13 (see the bottom part of the right panel of our Fig. 1). However, T+1 equations, for example, $A + 5 = G$, have an advantage in terms of rejection times over T-1 equations, for example, $A + 5 = E$, because the former contains the only combination between the letter augend and the proposed answer in the whole study set. This was also found by Zbrodoff, because her T+2 equations for +5 problems in day 13 were rejected faster than other equation types. Interestingly, the T+1 equations of the second-largest addend in our study, for example, $A + 4 = F$, have an advantage over T and T-1 equations of the same addend. This is probably because having memorized $A + 5 = F$ as true, participants could quickly reject $A + 4 = F$ as true.

As evoked above, the reversal of the sign of the difference between the two types of false equations for +4 and +5 problems was not found in the nonbreakers (see the top panel of Fig. 5), for whom solution times increased with addends for the three types of equations, that is, T, T-1, and T+1. Therefore, it seems likely that participants in this group mainly use counting until the end of practice, supporting the interpretation and conclusion of Thevenot

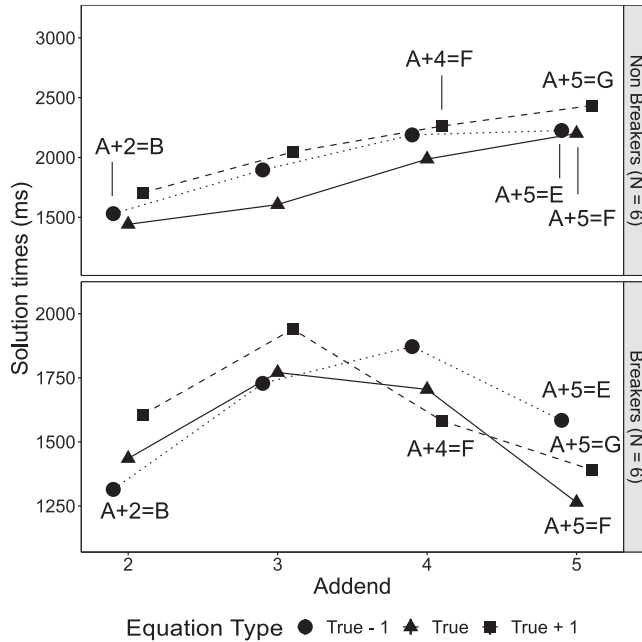


Fig 5. Illustration of reversal of the difference in rejection times for breakers and nonbreakers.

Note. Solution times as a function of addend and equation type (circles and dotted line for T-1, triangles and solid line for T, and squares dashed line for T+1 equations) in session 25 to illustrate the reversal of the sign of the difference in rejection times between within-count and outside-of-count equations for nonbreakers (top panel) and breakers (bottom panel).

et al. (2020). Thus, the results of the current paper reinforce those of Thevenot et al., namely that the possibility that counting was still used after an extensive practice cannot be discarded, and that retrieval is only used for a minority of problems and only for problems with the largest addend. Still, we have to keep in mind that opportunistic stopping might have been used by our participants but that the paradigm that we adopted from Zbrodoff (1990) is not suitable to reveal it. A possible reason for such a failure could be that in solving a within-count equation, when individuals have reached the proposed answer and opportunistically stopped counting, they have to judge it as false. This mismatch might induce a cognitive dissonance, and therefore the time gained by opportunistic stopping for false within-count equations could result in longer solution times than for true equations.

Nevertheless, several reasons could explain why participants do not take the opportunity to stop counting in alphabet-arithmetic tasks. An explanation can be found in Sternberg (1966) in his high-speed memory search experiment. Sternberg asked participants to memorize a sequence of 1–6 items (digits or letters) that constituted a memory set. He then showed a digit or letter that was either one of the items in the memory set or another item (i.e., a target or distractor). The participants had to decide whether this digit or letter was among the items in the memory set. The search for a distractor is necessarily exhaustive because all items

in the memory set have to be scanned before the “no” decision can be taken. In this case, response times should be a function of the size of the memory set. The search for a target, on the other hand, is not necessarily exhaustive because participants could stop scanning the memory set once the matched item is found. In this case, response times would be a function of the position of the item in the memory set. However, Sternberg found that response times for both the distractors and targets were a function of memory-set size with similar slopes, that is, 30 and 40 ms/item for the targets and distractors, respectively. Therefore, Sternberg concluded that individuals do not stop searching the memory set even after they have found a matching digit but scan the whole list before responding. The results of our experiment show that the same behavior is adopted by individuals when they have to take a decision involving a counting sequence. Another reason as to why participants do not opportunistically stop counting in alphabet-arithmetic tasks could be that they use a counting-up strategy from the letter augend to the letter answer. This strategy would consist of counting the number of letters separating the two letters given in the equation (e.g., for the equation $A + 5 = G$, counting from B to G results in six letters) and comparing this number of counts to the addend (6 is not 5, hence the equation is not true). This strategy does not provide participants with an opportunity to stop counting.

With some exceptions, in our current paper and in Zbrodoff (1999)’s Experiment 4, true equations were solved faster than false equations. This might be due to the fact that in both studies, true equations were presented twice as often as false ones. Another possibility is that, following the proposition of Ashcraft and Battaglia (1978) for mental arithmetic, participants first find the searched-for answer and then compare it with the proposed answer. The comparison time depends proportionally on the distance between the correct and proposed answer, such that true equations are solved faster than false equations. It is possible that participants in alphabet-arithmetic verification task adopted the same strategy.

All in all, our conclusions do not support the idea that opportunistic-stopping study could reveal a shift from counting to retrieval. Therefore, our results question the deduction of Zbrodoff (1999) and its support for the instance theory of automatization in particular (Logan, 1988). More generally, our results and those of Thevenot et al. (2020) question the implication of the instance theory of automatization on retrieval models of mental arithmetic (e.g., Ashcraft, 1982, 1992; Campbell, 1995; Campbell & Oliphant, 1992; Chen & Campbell, 2018; Siegler, 1996). In fact, the use of counting after an extensive practice could support the automated counting procedure theory (Barrouillet & Thevenot, 2013; Thevenot et al., 2020; Thevenot & Barrouillet, 2020; Uittenhove, Thevenot, & Barrouillet, 2016), according to which small additions with operands inferior to 5 are solved by adults through very fast counting procedures instead of retrieval. It is also possible that practice helps individuals to sharpen their understanding of number relations and to discover numerical patterns on which they can base their solution process through decompositions and derived-fact strategies (e.g., Baroody, 1985). Individuals could, therefore, memorize only a limited number of meaningful combinations and use them as a basis for developing their reasoning and solution processes, and even for inventing new solving strategies (e.g., Baroody, 1985; Baroody & Rosu, 2006). This kind of strategies would be particularly efficient when the operands in the problems are

too large for the implementation of a quick one-by-one counting procedure (Uittenhove et al., 2016).

More generally, even if addend slope in some of our participants show that counting can be the dominant strategy at the end of an intensive arithmetic training, this does not mean that counting is the unique strategy (Dewi & Thevenot, in revision). We have described the difference between nonbreakers and breakers and explain that the latter seemed to solve problems with the largest addends through memory retrieval. We have also explained how plausibility judgments can be implemented by individuals and how such judgments allow them to avoid counting (Table A.1). Moreover, it is possible that some participants used opportunistic stopping to solve some of the problems and that, as already evoked, either the paradigm that we used fail to reveal this strategy or that the sign of such incidental strategies is hidden in mean solution times. As also already stated, it is likewise possible that some participants memorized a limited number of combinations between letter addend, addend, and letter answer, which could constitute a basis for decomposing and procedurally processing other problems (e.g., knowing that $A + 2 = C$ could be used to solve $A + 3$, i.e., $A + 2 + 1$, hence D). Even in participants who showed a clear linear addend slope at the end of the experiment, we cannot exclude the possibility of infrequent use of retrieval for some problems. This variety of strategies in arithmetic has been often reported and analyzed in the literature (e.g., Bagnoud et al., 2021; LeFevre, Sadesky, & Bisanz, 1996; Siegler & Shrager, 1984) but we show here that retrieval might not be the dominant strategies, even after an extensive practice.

In terms of educational implications, our results are consistent with the view that the ultimate goal of primary instruction should be to foster number and operation sense and the meaningful memorization of basic sums by rote. Specifically, education should build on children's informal counting-based addition, encourage the discovery of patterns and relations, and use these arithmetic regularities to devise reasoning strategies (Henry & Brown, 2008; National Council of Teachers of Mathematics, 2000; National Mathematics Advisory Panel, 2008). The automatization of reasoning strategies provides an important basis for fluency with basic combinations, including transfer to unpracticed combinations (e.g., Baroody, 2006; Baroody, Bajwa, & Eiland, 2009; 2012; 2016). Such deep reasoning about numbers presupposes that numbers are accurately represented mentally and that children can easily navigate from one number to another without cognitive cost. Such assumptions provide an explanation as to why children's arithmetic skills are improved after interventions based on one-by-one counting on a number line or on fingers (e.g., Fuchs et al., 2010). In the conventional view, such training leads to better memorization of arithmetic facts than drill does.

Finally, our results can be used to sharpen computational and mathematical modeling and simulation of learning, which are at the heart of artificial intelligence. Our conclusion that practice of procedures could develop into automatized counting and could foster reasoning-based strategies allows for a conception of more complete learning models mainly based on attention and working memory. Indeed, these cognitive factors play a central role in procedural models of arithmetic because after encoding the problem elements, individuals need to sequentially execute a series of procedural steps while keeping a goal active in working memory (e.g., Aim = execution of 5 steps). Encoding, refreshing key elements in working memory, such as the number of steps already executed, the intermediary results reached during the

solving process and the goal itself, speed of working memory decay, and forgetting threshold are, therefore, crucial parameters that need to be integrated in the models (Chouteau, Mazens, Thevenot, Dewi, & Lemaire, 2021). The notion of competition or choice between counting and procedure algorithms on the one hand and direct memory retrieval on the other hand, which is reflected in our results, for example, by different patterns of solution time distributions in breakers and nonbreakers (see also Appendix B), should and will be part of our future research.

Acknowledgments

We would like to express our gratitude to one anonymous reviewer for providing us with a table that illustrates the different kinds of plausibility judgments and also for pointing out the possibility of using the counting-up strategy in solving alphabet-arithmetic equations. We would also like to thank another anonymous reviewer for her or his remark that opportunistic stopping could be accompanied by a cognitive dissonance that leads to within-count equations being solved slower than true equations.

Open Access Funding provided by Universite de Lausanne.

Data

The data and E-prime code (i.e., material and script) of the experiment are available on <https://osf.io/py6wr/>

References

- 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., & Battaglia, J. (1978). Cognitive arithmetic: Evidence for retrieval and decision processes in mental addition. *Journal of Experimental Psychology: Human Learning and Memory*, 4(5), 527–538. <https://doi.org/10.1037/0278-7393.4.5.527>
- Ashcraft, M. H., & Christy, K. S. (1995). The frequency of arithmetic facts in elementary texts: Addition and multiplication in grades 1–6. *Journal for Research in Mathematics Education*, 5, 396–421. <https://doi.org/10.2307/749430>
- Bagnoud, J., Dewi, J., Castel, C., Mathieu, R., & Thevenot, C. (2021). Developmental changes in size effects for simple 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., Dewi, J., & Thevenot, C. (2021). Differences in event-related potential (ERP) responses to small tie, non-tie and 1-problems in addition and multiplication. *Neuropsychologia*, 153, 107771. <https://doi.org/10.1016/j.neuropsychologia.2021.107771>
- Baroody, A. J. (1985). Mastery of basic number combinations: Internalization of relationships or facts? *Journal for Research in Mathematics Education*, 16(2), 83–98. <https://doi.org/10.2307/748366>

- Baroody, A. J. (1995). The role of the number-after rule in the invention of computational shortcuts. *Cognition and Instruction*, 13(2), 189–219. https://doi.org/10.1207/s1532690xci1302_2
- Baroody, A. J. (2006). Why children have difficulties mastering the basic number combinations and how to help them. *Teaching Children Mathematics*, 13(1), 22–31. <https://doi.org/10.5951/TCM.13.1.0022>
- 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>
- Baroody, A. J., Bajwa, N. P., & Eiland, M. (2009). Why can't Johnny remember the basic facts? *Developmental Disabilities Research Reviews*, 15(1), 69–79. <https://doi.org/10.1002/ddrr.45>
- Baroody, A. J., Eiland, M. D., Purpura, D. J., & Reid, E. E. (2012). Fostering at-risk kindergarten children's number sense. *Cognition and Instruction*, 30(4), 435–470. <https://doi.org/10.1080/07370008.2012.720152>
- Baroody, A. J., & Rosu, L. (2006). Adaptive expertise with basic addition and subtraction combinations: The number sense view. In A. J. Baroody & T. Torbeys (chairs), *Developing adaptive expertise in elementary school arithmetic. Symposium conducted at the Annual Meeting of the American Educational Research Association*.
- Baroody, A. J., Purpura, D. J., Eiland, M. D., Reid, E. E., & Paliwal, V. (2016). Does fostering reasoning strategies for relatively difficult basic combinations promote transfer by K-3 students? *Journal of Educational Psychology*, 108(4), 576–591. <https://doi.org/10.1037/edu0000067>
- 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. (1987). Network interference and mental multiplication. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(1), 109–123. <https://doi.org/10.1037/0278-7393.13.1.109>
- 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., & Graham, D. J. (1985). Mental multiplication skill: Structure, process, and acquisition. *Canadian Journal of Psychology*, 39(2), 338–366. <https://doi.org/10.1037/h0080065>
- 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). 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>
- Chen, Y., Orr, A., & Campbell, J. I. D. (2020). What is learned in procedural learning? The case of alphabet arithmetic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(6), 1165–1177. <https://doi.org/10.1037/xlm0000775>
- Chouteau, S., Mazens, K., Thevenot, C., Dewi, J., & Lemaire, B. (2021). A computational model of counting along a mental number line. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 43, 2010–2016.
- Compton, B. J., & Logan, G. D. (1991). The transition from algorithm to retrieval in memory-based theories of automaticity. *Memory & Cognition*, 19(2), 151–158. <https://doi.org/10.3758/BF03197111>
- Dewi, J. D. M., Bagnoud, J., & Thevenot, C. (2021). Do production and verification tasks in arithmetic rely on the same cognitive mechanisms? A test using alphabet arithmetic. *Quarterly Journal of Experimental Psychology*, 74(12), 2182–2192. <https://doi.org/10.1177/17470218211022635>
- Dewi, J., & Thevenot, C. (in revision). Individual differences in the evolution of counting strategy: The case of alphabet-arithmetic tasks.
- 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>
- Fuchs, L. S., Powell, S. R., Seethaler, P. M., Fuchs, D., Hamlett, C. L., Cirino, P. T., & Fletcher, J. M. (2010). A framework for remediating number combination deficits. *Exceptional Children*, 76(2), 135–156. <https://doi.org/10.1177/001440291007600201>
- Grabner, R. H., Brunner, C., Lorenz, V., Vogel, S. E., & De Smedt, B. (2021). Fact retrieval or compacted counting in arithmetic—A neurophysiological investigation of two hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <https://doi.org/10.1037/xlm0000982>

- Groen, G. J., & Parkman, J. M. (1972). A chronometric analysis of simple addition. *Psychological Review*, 79, 329–343. <https://doi.org/10.1037/h0032950>
- Henry, V. J., & Brown, R. S. (2008). First-grade basic facts: An investigation into teaching and learning of an accelerated, high-demand memorization standard. *Journal for Research in Mathematics Education*, 39(2), 153–183. <https://doi.org/10.2307/30034895>
- Jerman, M. (1970). A counting model for simple addition. *Educational Studies in Mathematics*, 2(4), 438–445. <https://doi.org/10.1007/BF00417626>
- Landauer, T. K. (1962). Rate of implicit speech. *Perceptual and Motor Skills*, 15, 646. <https://doi.org/10.2466/pms.1962.15.3.646>
- LeFevre, J.-A., Sadesky, G. S., & Bisanz, J. (1996). Selection of procedures in mental addition: Reassessing the problem size effect in adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(1), 216–230. <https://doi.org/10.1037/0278-7393.22.1.216>
- Lemaire, P., & Fayol, M. (1995). When plausibility judgments supersede fact retrieval: The example of the odd-even effect on product verification. *Memory & Cognition*, 23(1), 34–48. <https://doi.org/10.3758/BF03210555>
- Lemaire, P., & Reder, L. (1999). What affects strategy selection in arithmetic? The example of parity and five effects on product verification. *Memory & Cognition*, 27(2), 365–382. <https://doi.org/10.3758/BF03211420>
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95(4), 492–527. <https://doi.org/10.1037/0033-295X.95.4.492>
- Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(2), 179–195. <https://doi.org/10.1037/0278-7393.17.2.179>
- Masse, C., & Lemaire, P. (2001). Do people combine the parity- and five-rule checking strategies in product verification? *Psychological Research*, 65, 28–33. <https://doi.org/10.1007/s004260000030>
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics: An overview*. National Council of Teachers of Mathematics.
- National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. U.S. Department of Education.
- Parkman, J. M., & Groen, G. J. (1971). Temporal aspects of simple addition and comparison. *Journal of Experimental Psychology*, 89(2), 335–342. <https://doi.org/10.1037/h0031198>
- Reder, L. M. (1982). Plausibility judgments versus fact retrieval: Alternative strategies for sentence verification. *Psychological Review*, 89(3), 250–280. <https://psycnet.apa.org/doi/10.1037/0033-295X.89.3.250>
- Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. Oxford University Press.
- Siegler, R. S., & Jenkins, E. (1989). *How children discover new strategies*. Lawrence Erlbaum.
- Siegler, R. S., & 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). Lawrence 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). Lawrence Erlbaum.
- Sternberg, S. (1966). High-speed scanning in human memory. *Science*, 153(3736), 652–654. <https://doi.org/10.1126/science.153.3736.652>
- 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., 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, 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>

Wenger, M. (1999). On the whats and hows of retrieval in the acquisition of a simple skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(5), 1137–1160. <https://doi.org/10.1037/0278-7393.25.5.1137>

Wilkins, N. J., & Rawson, K. A. (2011). Controlling retrieval during practice: Implications for memory-based theories of automaticity. *Journal of Memory and Language*, 65(2), 208–221. <https://doi.org/10.1016/j.jml.2011.03.006>

Zbrodoff, N. J. (1995). Why is 9 + 7 harder than 2 + 3? Strength and interference as explanations of the problem-size effect. *Memory and Cognition*, 23(6), 689–700. <https://doi.org/10.3758/BF03200922>

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(2), 299–317. <https://doi.org/10.1037/0278-7393.25.2.299>

Zbrodoff, N. J., & Logan, G. D. (1990). On the relation between production and verification tasks in the psychology of simple arithmetic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(1), 83–97. <https://doi.org/10.1037/0278-7393.16.1.83>

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.

Appendix A

Plausibility Judgments

Table A.1
Items affected by rule-based plausibility judgments in two studies

		Example: answers to letter augend A Possible addends: 3–5			
Type		A + 3	A + 4	A + 5	
Zbrodoff (1999)	T – 2	B = Letter After A	C = Skip Letter After A	D	
	T – 1	C = Skip Letter After A	D	E	
	T	D	E	F	
	T + 1	E	F	G	
	T + 2	F	G	H	
		Example: answers to letter augend A Possible addends: 2–5			
Type		A + 2	A + 3	A + 4	A + 5
The present study	T – 1	B = Letter After A	C = Skip letter After A	D	E
	T	C	D	E	F
	T + 1	D	E	F	G

Appendix B

Performance of Breakers and Nonbreakers

Figures B.1 and B.2

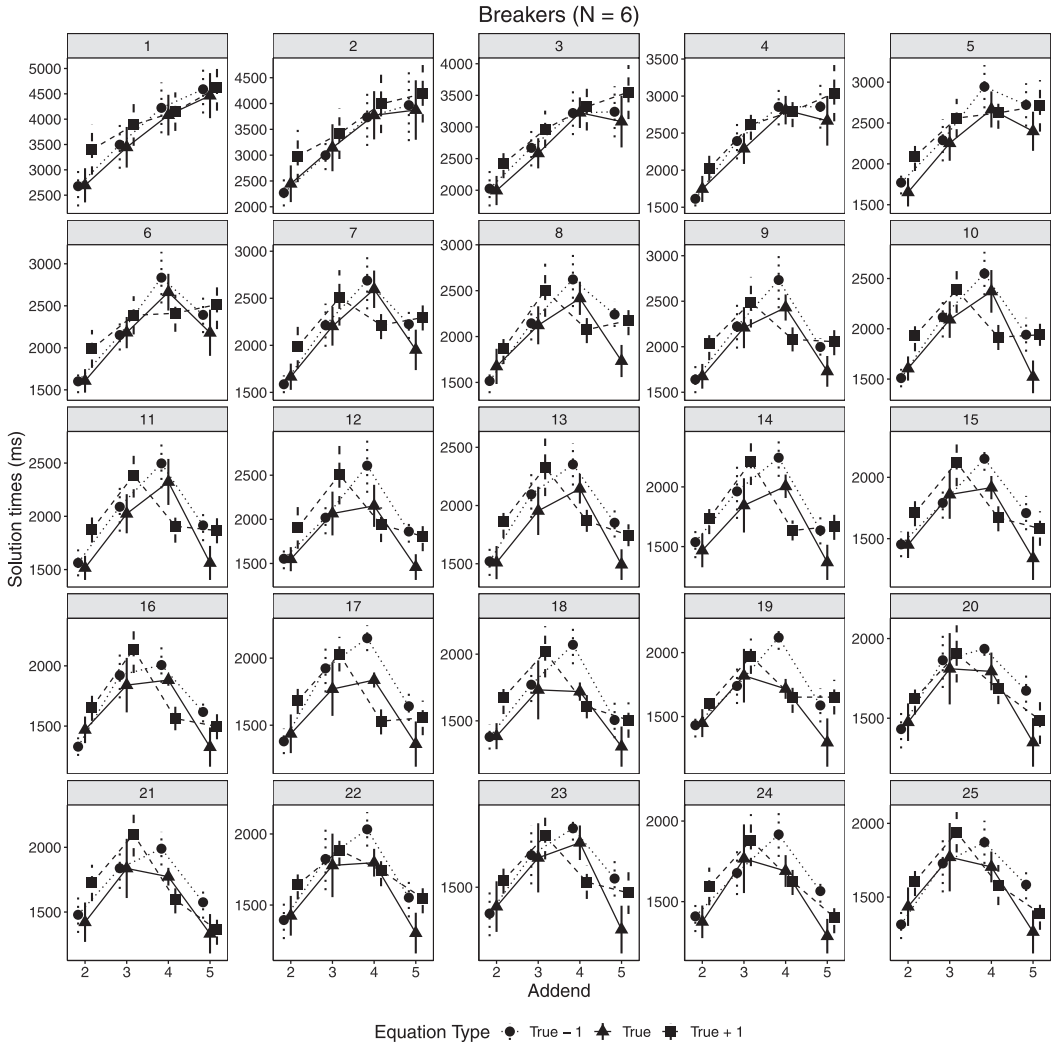


Fig B.1. Performance of breakers during 25 sessions.

Note. Solution times as a function of addend and equation type (circles and dotted lines for T-1, triangles and solid lines for T, and squares dashed lines for T+1 equations) for the breakers in 25 sessions. Error bars represent standard errors.

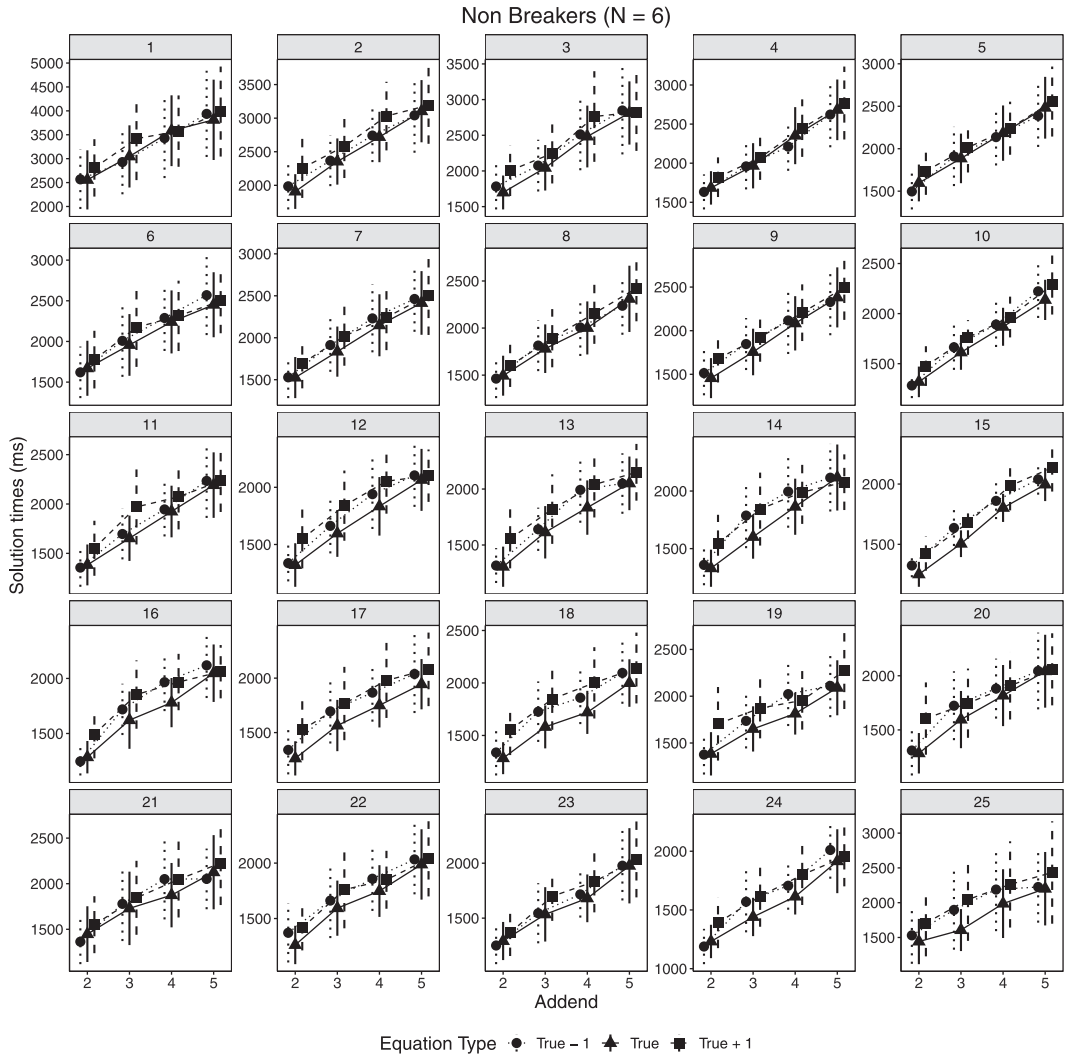


Fig B.2. Performance of nonbreakers during 25 sessions.

Note. Solution times as a function of addend and equation type (circles and dotted lines for T-1, triangles and solid lines for T, and squares dashed lines for T+1 equations) for the nonbreakers in 25 sessions. Error bars represent standard errors.