Memory & Cognition (2021) 49:1236–1246
https://doi.org/10.3758/s13421-021-01161-6

Effects of prior-task failure on arithmetic performance: A study in young and older adults

Patrick Lemaire[1]

Accepted: 19 February 2021 / Published online: 8 March 2021
© The Psychonomic Society, Inc. 2021

Abstract
Effects of prior-task failure (i.e., decreased performance on a target task following failure on a prior task) were tested in young and older adults. Young and older participants (N=120) accomplished a computational estimation task (i.e., providing the best estimates to arithmetic problems) before and after accomplishing a dot comparison task in a control or in a failure condition. Both groups decreased their performance on the target computational estimation following failure on the prior dot comparison task. Also, prior-task failure led young and older adults to select the better strategy less often and to use the easier strategy more often. Our findings show, for the first time, impaired performance after experiencing failure in both young and older adults. We discuss implications of these findings for further our understanding of effects of task transitions (i.e., prior-task success and failure) on cognitive performance.

Keywords Prior-task failure · Arithmetic · Cognitive aging · Strategies · Effects of task transition

Introduction

The goal of this study was to compare effects of prior-task failure on young and older adults' arithmetic performance. Effects of prior-task failure refer to decreased performance on a target task following failure on a prior task (Smith, Kass, Rotunda, & Schneider, 2006). These effects have been evidenced in young adults only. Thus, we do not know if they occur in older adults, and if they do whether young and older adults are differently sensitive to experiencing failure on a prior task. Also, the mechanisms responsible for these effects are unclear. The present study aimed at addressing these issues. Before presenting the logic of the present experiment, I discuss why it is important to further our understanding of effects of prior-task failure and review previous findings on these effects.

Effects of prior-task failure are important to study for several reasons. First, because they are among many situational factors (e.g., speed/accuracy pressure, social comparisons, stressful environment) that influence cognitive performance, they potentially inform how such situational factors greatly affect participants' cognitive performance, both in lab experiments for testing theories of cognition and in clinical settings during neuropsychological assessments. Second, a good understanding of effects of prior-task failure may fruitfully inform age-cognition relations. These relations are importantly modulated by a variety of situational factors, such as age-based stereotype threat (see Barber & Lui, 2020, for a recent overview), intergenerational interactions (e.g., Abrams et al., 2008), or induced subjective age (e.g., Haslam et al., 2012). Like other situational factors, effects of prior-task failure may change age-related differences in cognitive performance in several different ways. Thus, it is possible that experiencing failure on a given task increases age differences in performance on a subsequent task if older adults are more sensitive to prior-task failure than young adults and older adults' performance on a subsequent task declines because of this task failure. Alternatively, age differences in a subsequent cognitive task may decrease if prior-task failure influences older adults less than young adults, and young adults' performance decreases more than older adults' following prior-task failure. Moreover, it is also possible that, for some cognitive domains, like arithmetic tested here, where baseline age-related differences are either much smaller or even non-existent (see Uittenhove & Lemaire, 2015, 2018, for overviews), age differences in sensitivity to effects of prior-task failure would result in age differences in participants' performance, independently of age differences in cognitive capacities.

✉ Patrick Lemaire
patrick.lemaire@univ-amu.fr

[1] LPC, and CNRS, Aix-Marseille Université, 3 Place Victor Hugo, Case D, 13331 Marseille, France

Modulations of age-related differences in cognitive performance by prior-task failure would have important implications on theories of cognitive aging in general and for thinking of older adults' cognitive capacities in particular. Indeed, if prior-task failure increases age-related differences, this would imply that some situational factors might limit older adults in using their potential resources and/or amplify the deleterious consequences of decreased processing resources with age.

Although they can be viewed as analogs of effects of negative feedback found many times (e.g., Butler & Roediger, 2008; Strickland-Hughes, West, Smith, & Ebner, 2017; West, Dark-Freudeman, & Bagwell, 2009), effects of prior-task failure and of feedback are different in many respects. For example, effects of prior-task failure can be observed with no feedback provided by experimenters (e.g., Smith et al., 2006). As another example, when participants experience feedback, it is based on their actual performance in prior-task failure experiments (e.g., Geraci & Miller, 2013), whereas it is not necessarily based on actual performance in feedback experiments (e.g., Strickland-Hughes et al., 2017). In other words, the role of feedback per se is not of interest in effects of prior-task failure. Feedback is sometimes provided to make sure that participants know how much they actually failed in the failure condition.

Smith et al. (2006) were the first to report effects of prior-task failure. They asked young adults to solve anagrams. During pretest and post-test, participants completed ten average-difficulty anagrams. Before post-test, participants were asked to attempt five unsolvable anagrams and five difficult anagrams in the failure condition or ten average-difficulty anagrams in the control condition. Both the control and failure groups performed equally well at pretest (they succeeded on 70% of anagrams). However, at post-test, the failure group performed much more poorly than the control group (55% vs. 66%). Thus, experiencing failure led participants to perform more poorly.

Effects of prior-task failure are not always found. For example, Geraci and Miller (2013) asked participants to accomplish a verbal free-recall task. Before the memory test, participants were randomly assigned to one of three experimental groups. In the first group, participants were given 30 sets of five scrambled words. For each set of five words, participants were asked to rearrange the words to form a grammatically correct four-word sentence. This task was successfully accomplished by participants in the unlimited-time success condition. However, this sentence-scramble task was not successfully accomplished when participants had only 10 s/sentence (failure condition). Finally, in the control group, participants had no task prior to the target memory task. Geraci and Miller found that older, but not young, participants correctly recalled more words in the success condition than in the control or failure conditions. As a result, young and older adults performed equally well in this success condition. Also, Geraci and Miller found no effects of prior-task failure, as there were no differences between control and failure conditions in both young and older adults, and older adults obtained poorer memory performance than young adults in both the control and failure conditions.

It is possible that manipulation of failure was not strong enough in Geraci and Miller's (2013) study. It is also possible that effects of prior-task failure are specific to the context (i.e., anagram problem solving) tested by Smith et al. and would not occur in other contexts (like episodic memory, as tested by Geraci and Miller). Finally, it is also possible that the effects of prior-task failure reported by Smith et al. (2006) were just mere sequential difficulty effects. Indeed, participants in the control condition of Smith et al.'s study solved easier, solvable anagrams in contrast to participants in the failure condition who solved more difficult and unsolvable anagrams. Difficult and unsolvable anagrams of the failure condition likely require more cognitive resources than easier anagrams in the control condition. Fewer available resources during post-test in the failure condition may have led to poorer performance relative to the control condition, independently of success or failure. Although such cross-task sequential difficulty effects have never been found, sequential difficulty effects have already been found many times across items (e.g., Schneider & Anderson, 2010; Uittenhove & Lemaire, 2013a, 2013b). To test effects of prior-task failure, it is important to equate as much as possible prior-task difficulty, such that effects of prior-task failure are not confounded by sequential difficulty effects. In the present study, we used a dot comparison task in which participants are presented two dot collections and have to determine the more numerous dot collection. This task was used because it is very easy (i.e., participants take less than 1 s on each trial; e.g., Halberda, Ly, Wilmer, Naiman, & Germine, 2012) and yet participants can fail when they compare two dot collections with the same number of dots.

The present study aimed at testing more strongly effects of prior-task failure and at determining their conditions of occurrence and their underlying mechanisms, as well as whether they differ in young and older adults. Previous research found that older adults are more sensitive than young adults to effects of prior-task success (see Geraci & Miller, 2013; Geraci et al., 2016; Lemaire & Brun, 2018; Lemaire, Gouraud, & Nicolas, 2019), such that experiencing success in one task improved older adults' performance on a subsequent task more than young adults'. In contrast, previous studies revealed that effects of prior-task failure were found in young adults (Smith et al., 2006) but not in older adults (e.g., Geraci & Miller, 2013). It is surprising that older adults' cognitive performance on a target task was not influenced by their failure on an immediately preceding task in previous studies (e.g., Geraci & Miller, 2013). Indeed, given age-related decrease in cognitive performance and given how older adults are sometimes more sensitive than young adults to situational or contextual factors (e.g., speed pressures; stereotype threat; stress; framing), we could expect that older adults suffer more from prior-task failure than young adults. For example, experiencing failure might reinforce an age-based stereotype threat, known to impair older adults' cognitive performance and confidence (e.g., Nicolas et al., 2020; see Barber, 2017; Lamont, Swift, & Abrams, 2015, for reviews). Note though that it is also possible that older adults are truly not influenced by prior-task failure. This could occur via several non-exclusive mechanisms (e.g., negative information discounting, negative emotion regulation strategies). Thus, older adults may counteract negative feelings associated with prior-task failure and/or not take into account their experience of failure on a prior task. This would lead them to approach the target task independently of what happened on the immediately preceding task. The goal of the present study was to test more strongly effects of prior-task failure and how these effects differ in young and older adults. Replicating no effects of prior-task failure in older adults (as suggested by Geraci & Miller's findings) and effects of prior-task failure in young adults (as found by Smith et al.) would show that young and older adults differ in their sensitivity to prior-task failure. If that is the case, we would have to find out how this occurs. At a more general and empirical level, this would mean that effects of task transition (i.e., prior-task success or failure) vary with aging, such that young adults would be most sensitive to prior-task failure and older adults to prior-task success. Alternatively, finding larger effects of prior-task failure in older than in young adults would suggest that age-related differences in cognitive performance in one task may be enhanced by prior-task failure. This would be another manifestation of cognitive aging being modulated by psychosocial, situational factors.

To determine via which mechanisms prior-task failure influences participants' performance and whether effects of prior-task failure occur via the same mechanisms in young and older adults, we adopted a strategy approach (Lemaire, 2010, 2016). As a strategy is defined as a "procedure or a set of procedures to achieve a higher-level goal" (Lemaire & Reder, 1999, p. 365), a strategy approach enables knowing whether prior-task failure changes the way young and older adults accomplish cognitive tasks. The strategy approach is based on previous findings showing that condition-related and age-related differences in cognitive performance are mediated by strategic variations. In short, participants do not always use the available strategies equally often, and do not execute or select strategies with comparable levels of efficacy to accomplish cognitive tasks in different conditions (see Lemaire, 2016, for a review). In the present study, we tested the hypothesis that effects of prior-task failure impair participants' performance via selection of the less efficient strategy on each problem and/or via poorer execution of the selected strategies. To test this hypothesis, in a pretest–post-test design, young and older adults were given a target, arithmetic problem-solving task before and after experiencing task failure in a prior-task failure condition. Their performance was compared to participants' performance in a control condition where participants did not experience prior-task failure.

We tested effects of prior-task failure in young and older adults in the context of arithmetic for a couple of reasons. First, to further our knowledge of mechanisms underlying arithmetic performance, it is important to determine how situational factors, like prior-task failure, influence arithmetic performance. Arithmetic performance is known to vary with a number of factors, including problems and participants' characteristics, as well as task environment and situational parameters (see Cohen Kadosh & Dowker, 2015; Gilmore, Göbel, & Inglis, 2018, for overviews). Several situational parameters have already been established to crucially affect participants' performance in arithmetic, like prior-task success (Lemaire & Brun, 2018; Lemaire et al., 2019), stereotype threat (e.g., Beilock, 2008; Beilock & DeCaro, 2007), speed-accuracy pressures (e.g., Lemaire, Arnaud, & Lecacheur, 2004), response deadlines (e.g., Campbell & Austin, 2002), or problem formats (e.g., Mauro, LeFevre, & Morris, 2003). Second, arithmetic is one of the few cognitive domains where age-related differences are either non-existent or much smaller than in other cognitive domains (Uittenhove & Lemaire, 2015, 2018). Controlling for comparable baseline performance between young and older adults is important when testing age-related differences in effects of prior-task failure because larger effects of prior-task failure in older adults, for example, may be contaminated by (or even the result of) older adults' lower baseline performance.

In the present study, we asked young and older adults to accomplish a computational estimation task (i.e., estimating products of two-digit multiplication problems like 48 × 72). Computational estimation tasks bear important similarities to most other cognitive tasks (i.e., accuracy and latency performance, and variations in these as a function of item and participant characteristics, can be assessed). This enables generalization of the present findings outside the context of this task. Moreover, direct (rather than indirect) measures of which strategy is used on each item are easily collected in computational estimation tasks, given available external evidence (such as when participants are calculating out loud, as they did here). Such measures are independent of participants' performance and are fruitful to provide a mechanistic account of effects of prior-task failure. Moreover, the better strategy (i.e., the strategy among available strategies that yields the approximate product that is closest to the correct product) for a given problem is very clear in computational estimation tasks. For example, in a situation like here where participants can choose between a rounding-down strategy (or rounding both operands down to the nearest smaller decades like doing 30 × 60

= 1,800 to estimate 34 × 63) or a rounding-up strategy (or rounding both operands up to the nearest larger decades like doing 30 × 60 = 1,800 to estimate 27 × 59), it is easy to know which strategy is the better strategy on each problem. Thus, rounding down is a better strategy for problems like 34 × 63 or 32 × 56, and rounding up is a better strategy for problems like 27 × 59 or 34 × 48. Finally, previous works (e.g., Lemaire et al., 2004) found that older adults suffer much less, if any, performance decline, in this task – a nice feature to determine if young and older adults show similar or different effects of prior-task failure.

Participants were tested either under a prior-task control condition or a prior-task failure condition. We assessed which strategy was used and participants' performance on each problem. We compared young and older adults' strategy use and performance before and after they took a dot-comparison task that participants accomplished either unsuccessfully or much more successfully. These data enabled us to test effects of prior-task success (i.e., lower performance following prior-task failure relative to following no prior-task failure) and the strategy hypothesis. The strategy hypothesis states that changes in how often participants select the better strategy and/or in how they execute strategies are responsible for effects of prior-task failure. This strategy hypothesis predicts decreased use of the better strategy in the post-test relative to pretest for participants tested under the failure condition. These effects were expected above and beyond mere test-retest effects found in the control condition.

Regarding age-related differences in effects of prior-task failure, the following hypotheses were tested. First, like effects of prior-task success, only older adults would be influenced by failure on a prior task when accomplishing a target task. This predicts that older adults' performance, but not that of young adults, would decrease on a target task after experiencing failure on a prior task. This would suggest that only older adults are sensitive to prior-task performance when they accomplish a target task, such that success on a prior task improves their performance on a target task and failure impairs this performance. A variant of this hypothesis is that both age groups show effects of prior-task failure, but these effects are stronger in older adults. This predicts that decreased performance following failure would be larger in older than in young adults. Such findings would occur if older adults are more sensitive than young adults to poor performance on a prior task when they accomplish a target task. Alternatively, following Geraci and Miller (2013), who found no effects of prior-task failure in either young or older adults, and following Smith et al. (2006), who found effects of prior-task failure in young adults, older adults might be less sensitive than young adults to effects of prior-task failure. This could occur if older adults discount prior-task failure, feel less strongly about their failure, or better regulate negative feelings resulting from prior-task failure.

In addition to investigating age-related differences in effects of prior-task failure, by testing the strategy hypothesis, the present study enabled us to determine whether effects of prior-task failure occur via the same or different mechanisms in young and older adults. Finding that prior-task failure leads both young and older participants to use the better strategy on each problem less often would suggest that effects of prior-task failure result from similar mechanisms in both age groups. In contrast, finding that prior-task changes strategy use in only one group would imply that different mechanisms are responsible for effects of prior-task failure in young and older adults.

Method

Participants A total of 120 participants were tested: 60 young and 60 older adults (see Table 1 for participants' characteristics). Half the participants were randomly assigned to the control condition, and half to the failure condition. Our target sample size was determined using an *a priori* power analysis (G*Power; Faul, Erdfelder, Lang, & Buchner, 2007). Although no previous studies tested age-related differences in effects of prior-task failure, we used η^2_p from previous studies on prior-task success. Previous results on prior-task success in young and older adults found that η^2_p ranged from .05 to .12 (Geraci et al., 2016; Geraci & Miller, 2013). Using a η^2_p =.05, our study design of two between-participants factors (age and condition) and one repeated factor (testing), could achieve 80% power with 56 participants. In order to exceed this criterion and achieve larger than 80% power, we recruited 120 participants.

Participants provided written informed consent. This study did not receive approval from a research ethics committee as it was not applied for – this is because, in France, we need such approval for brain-imaging (ERP, MEG, fMRI) data only; we do not need it for behavioral research. Nevertheless, the protocol has been in accordance with the ethical standards laid down in the Declaration of Helsinki. The author has no ethical conflicts or conflicts of interest to disclose.

Prior to the experiment, all older adults completed the Mini Mental-State Examination (MMSE; Folstein, Folstein, & McHugh, 1975). No older adults obtained scores lower than the usual cut-off score of 27; therefore, none were excluded.

Next, participants completed the French Kit (a pencil-and-paper arithmetic fluency test; French, Price, & Akstrom, 1963). In this test, participants had to correctly solve three subsets of basic arithmetic problems (i.e., addition, subtraction, and multiplication) in a total of 6 min. Each subset of basic arithmetic problems was presented for 2 min, and participants were asked to solve as many problems as possible within the limited duration. The number of correct answers on each subset was summed to yield a total arithmetic fluency

Table 1 Participants' characteristics

Characteristics	Young adults (N = 60)			Older adults (N = 60)			Age × Condition
	Control condition	Failure condition	F	Control condition	Failure condition	F	F
N (females)	30 (22)	30 (23)	---	30 (19)	30 (16)	---	---
Age (SD)	21.5 (1.82)	20.2 (1.62)	1.43	72.0 (5.41)	71.1 (5.19)	0.35	0.24
Range	(19—24)	(18—25)	---	(65—83)	(68—83)	---	---
French Kit (SD)	33 (17.06)	31 (12.67)	0.33	57 (17.22)	55 (13.88)	1.02	1.93
MHVS[1] (SD)	18 (4.60)	20 (3.74)	2.09	25 (4.16)	25 (4.52)	0.02	0.48
MMSE[2]	---	---	---	30 (0.48)	30 (0.55)	0.76	---

[1] Mill-Hill Vocabulary Scale

[2] Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975)

** p<.01

score. As is often found, older adults had significantly larger arithmetic fluency scores than young adults, $F(1,118)=34.89$, $MSe=232.7$; $\eta^2_p=.23$.

Then, to assess individuals' verbal fluency, participants completed the French version of the Mill–Hill Vocabulary Scale (MHVS; Deltour, 1993; Raven, Raven, & Court, 1998). MHVS consists of 33 items distributed across two pages. Each item was a target word followed by six proposed words, and the task consisted of identifying which of the proposed words had the same meaning as the target word. The number of correct items represented the level of individuals' verbal ability. As is often found in aging research, young adults obtained poorer verbal fluency scores than older adults, $F(1,118)=45.67$, $MSe=17.53$; η^2_p =.28. Note that within each age group, participants tested in the failure and control conditions did not differ (Fs<1.11).

Stimuli for the prior dot comparison task Stimuli were collections of black dots displayed on a white background. More specifically, in each trial, two collections of dots were presented side-by-side on a laptop and varied in number (i.e., one dot array always represented 24 dots, and the other 18, 20, 22, 26, 28, or 30 dots). Dots were randomly distributed and were at least 0.6 cm (or 25 pixels) away from each other to avoid dots overlap.

Stimuli were selected on the basis of participants' success rates in a previous study (Roquet & Lemaire, 2019). Thus, collections of dots for which more than 90% of participants in Roquet and Lemaire's study gave an incorrect response were tested in the failure condition. Collections of dots for which 50% of participants gave a correct response were tested in the control condition. This led participants to have more chances to experience failure in the failure condition.

Stimuli for the target computational estimation task Based on previous works (e.g., Hinault, Badier, Baillet, & Lemaire, 2017; Lemaire & Brun, 2016; Lemaire & Leclère, 2014), the computational estimation task included 42 two-digit multiplication problems (e.g., 34 × 67). Each problem had the unit digit of one of the operands smaller than 5 and the unit digit of the other operand larger than 5 (e.g., 43 × 69). In half the problems, the sum of unit digits was smaller than 10 so that they were better estimated with the rounding-down strategy. In the other problems, the sum of unit digits was larger than 10, so that they were better estimated with the rounding-up strategy. We matched correct sums and mean percent deviations for problems that were best estimated with the rounding-down strategy and for problems that were best estimated with the rounding-up strategy, so that strategy selection would not be contaminated by these factors.

Following previous findings in arithmetic (see Kadosh & Dowker, 2015, for an overview), problems were selected with the following constraints: (a) no operands had a 0 unit digit (e.g., 20 × 63) or a 5 unit digit (e.g., 25 × 63); (b) no digits were repeated within operands (e.g., 22 × 63); (c) no reverse orders of operands were used (e.g., 24 × 63 and 63 × 24); (d) the first operand was larger than the second operand in half the problems, and vice versa; (e) no operand had its closest decade equal to 0, 10, or 100; and (f) problems were randomly presented with the constraint that rounded operands were never the same across two consecutive rounding problems in a given trial (e.g., if one problem was 32 × 64, the next problem could not be 31 × 62).

Procedure The E-Prime software-controlled stimulus display and latency collection for both the dot comparison and the computational estimation tasks. The experiment was conducted in one session that lasted nearly 60–75 min (*mean*: 68 min; *SD*=10.5).

First, older adults took the MMSE. Then, both young and older adults took the Mill-Hill Vocabulary test and the French Kit.

Next, young and older adults accomplished the computational estimation task. Two-digit multiplication problems were presented horizontally in 84-pt bold black Courier font in the middle of a 15-in. white computer screen. Participants were asked to provide estimates to these problems using either a rounding-up or a rounding-down strategy. The rounding-down strategy was described as rounding both operands down to the nearest decades, for instance doing 40 × 30 = 1,200 to estimate 41 × 36, and the rounding-up strategy as rounding both operands up to the nearest decades, for instance doing 40 × 20 = 800 to estimate 34 × 19. All participants completed a short training and practiced the task on eight problems to familiarize themselves with the procedure and the task. Participants were explicitly told that, for some problems, the better strategy involved rounding one operand down and the other up to the closest decades (e.g., 34 × 19), but that this mixed-rounding strategy was not allowed. This mixed-rounding strategy was not allowed to make the strategy choice process harder, given that previous studies showed that when participants can use mixed-rounding, strategy selection is so easy that everybody selects the better strategy on more than 95% of problems (e.g., Lemaire et al., 2004). Participants used no other strategies than rounding down or rounding up during both pretest and post-test.

Each trial started with a 300-ms blank screen before a 400-ms fixation cross displayed at the center of the screen, followed by one problem. Following numerous previous works using this procedure (e.g., Lemaire & Leclère, 2014), the timing of each response began when the problem appeared on the screen and ended when the experimenter pressed the right button of a two-button mouse, the latter event occurring as soon as possible after participants' responses. Participants were asked to calculate out loud to determine which strategy they used. On each problem, the experimenter wrote down participants' estimation responses and strategy choice.

Following this pretest of computational estimation, participants performed the prior (i.e., dot comparison) task. On each trial, participants saw two arrays of dots for 1,500 ms and had to indicate, as quickly and accurately as possible, which collection was most numerous. In both the failure and the control conditions, participants received feedback twice: (1) first, success or failure feedback (i.e., with a green square for success and red square for failure) after each trial, and (2) second, a success rate with the message "You erred on X% of trials" at the end of the task (X being equal to the participant's failure rate).

Next, all participants accomplished the same computational estimation task as in the pretest with the following differences: (a) the order of operands for each problem was reversed (e.g., if 78 × 23 was tested during pre-test, 23 × 78 was tested during post-test), (b) the order of problems was random but different from that of the order of problems during pretest. In both pretest and post-test, no feedback was provided on the computational estimation task performance. All participants had a short break in the middle of the computational estimation task (after 21 problems), and before and after prior-dot comparison tasks.

Results

Results are reported in three main sections. First, participants' performance (mean percent error rates and response times) on our prior, numerosity judgment task was analyzed. Then, participants' performance on our target computational estimation task was examined to determine whether young and older adults' performance decreased between pretest and post-test more in the failure than in the control condition. Third, how prior-task failure influenced strategy use (i.e., use of the better strategy and of rounding-down on each problem) was analyzed. Better strategy selection on a given problem was coded 1 if participants used the better strategy (i.e., rounding-down for rounding-down problems and rounding-up for rounding-up problems) and 0 otherwise (i.e., rounding-down for rounding-up problems and rounding-up for rounding-down problems). Numerosity judgment performance was analyzed with 2 (Age: young, older adults) × 2 (condition: control, failure) between-participants ANOVAs. Performance in computational estimation task (i.e., mean estimation latencies and absolute percentages of deviations between estimates and correct products) and strategy use (mean percentages of better strategy use and of rounding-down strategy use) were analyzed with mixed-design ANOVAs, 2 (Age: young, older adults) × 2 (Condition: control, failure) × 2 (Testing: pretest, post-test), with repeated measures on the last factor (see means in Table 2). Effects of prior-task failure were further examined with separate analyses in each age and condition group, followed by pretest–post-test pairwise comparisons. All the analyses were run with and without arithmetic fluency scores as covariates, as older adults' arithmetic fluency was significantly larger than that of young adults. Analyses with and without arithmetic fluency as covariates yielded exactly the same effects. In all results, unless otherwise noted, differences are significant to at least $p<.05$.

Performance in the prior numerosity judgment task As can be seen in Table 2, participants erred more often in the failure condition than in the control condition (77.5% vs. 44.8%; $F(1,116)=421.348$, $p<.001$, $MSe=76.2$, $\eta^2_p=.78$), and this was found in both young ($F(1,58)=127.854$, $p<.001$, $MSe=11482.6$, $\eta^2_p=.69$) and older adults ($F(1,58)= 341.593$, $p<.001$, $MSe=21407.4$, $\eta^2_p=.85$). Also, young adults erred more than older adults (62.9% vs. 59.3%; $F(1,116)=5.130$, $p=.025$, $MSe=76.2$, $\eta^2_p=.04$). The Age × Condition interaction was significant, $F(1,116)=10.055$, $p=.002$, $MSe=76.2$, $\eta^2_p=.08$, as the differences between failure and control conditions were larger in older (37.8%) than in young (27.7%) adults. This was the result of young adults erring more than

Table 2 Performance (percent errors, response times in ms) in prior dot comparison task, performance (mean estimation times, absolute percentages of deviations), and strategy use (percentages of better strategy use and rounding-down strategy) during pretest and post-test in the computational estimation task for young and older adults tested under the control or success condition

Testing	Young adults			Older adults		
	Control	Failure	*Means*	Control	Failure	*Means*
			Numerosity judgment task performance			
% Errors	49.1	76.8	62.9	40.4	78.2	59.3
RTs	1,130	1,084	1,107	1,204	1,016	1,110
			Percentages of deviation			
Pretest	18.6	17.1	17.8	15.6	15.4	15.5
Post-test	17.0	17.5	17.2	15.8	15.8	15.8
Means	17.8	17.3	17.5	15.7	15.6	15.6
Differences	-1.6**	0.4	-0.6†	0.1	0.4*	0.3*
			Estimation latencies (in ms)			
Pretest	8,527	8,296	8,411	7,424	8,680	8,052
Post-test	6,858	6,369	6,613	6,592	8,085	7,339
Means	7,692	7,332	7,512	7,008	8,383	7,695
Differences	-1,669**	-1,927**	-1,798**	-832**	-595*	-713**
			Better strategy selection (in %)			
Pretest	72.7	75.0	73.8	81.3	82.7	82.0
Post-test	76.4	72.0	74.2	83.5	78.2	80.8
Means	74.6	73.5	74.0	82.4	80.4	81.4
Differences	3.7**	-3.0*	0.4	2.2	-4.5*	-1.2
			Rounding-down strategy (in %)			
Pretest	54.4	58.5	56.4	54.0	62.0	58.0
Post-test	59.2	64.3	61.8	54.2	66.4	60.3
Means	56.8	61.4	59.1	54.1	64.2	59.1
Differences	4.8†	5.8*	5.3*	0.2	4.4*	2.3†

Note. ** p<.01; * p<.05, † p <.10

Differences: Post-test – Pretest

older adults in the control condition (49.1% vs. 40.4%; F(1,116)=14.774, p<.001, MSe=76.2, η^2_p=.11), and of both age groups erring equally often in the failure condition (76.8% vs. 78.2%; F<1.0). Also, participants were slower in the control condition than in the failure condition (1,160 ms vs. 1,050 ms; F(1,116)= 421.348, p=.004, MSe=76.2, η^2_p=.78). Thus, before accomplishing the computational estimation task in post-test, participants did fail much more often in the failure condition than in the control condition. No other effects showed significance on either error rates or response times.

Effects of prior-task failure on estimation performance

Percentages of deviation The following main and interaction effects were significant: Age, (F(1,116)=14.612, p<.001, MSe=14.81, η^2_p=.12), Age × Testing (F(1,116)=6.959, p<.001, MSe=11.34, η^2_p=.06), Condition × Testing interaction (F(1,116)=12.089, p<.001, MSe=1.63, η^2_p=.09), and Age × Condition × Testing interactions were significant (F(1,116)=7.232, p=.008, MSe=1.63, η^2_p=.06). Younger adults provided better estimates during post-test than during pretest in the control condition, F(1,58)=14.092, p<.001, MSe=39.136, η^2_p=.20, but equally good estimates across pretest and post-test (F<1.0) in the failure condition.[1] Older adults provided equally good estimates during pretest and post-test in the control condition (F<1.0), but poorer estimates during post-test than during pretest in the failure condition (F(1,58)=4.986, p=.03, MSe=2.401, η^2_p=.06).

Estimation latencies Two effects were significant: Testing, F(1,116)=67.245, MSe=1406591, η^2_p=.37; and Age × Testing, F(1,116)=12.551, MSe=1406591, η^2_p=.10. Participants were faster during post-test than during pretest, and this increased speed during post-test was larger in young adults (1,798 ms; F(1,58)=59.773, p<.001, MSe=1622541; η^2_p=.51) than in older adults (713 ms; F(1,58)=12.813, p<.001, MSe=1190641; η^2_p=.18). As can be seen in Table 2, increased speed between pretest and post-test was found for both young and older adults under both the control and failure conditions. No other effects were significant on estimation performance.

In summary, clear effects of prior-task failure were found on percentages of deviations in both young and older adults. Participants provided either equally good or better estimates during post-test than during pretest in the control condition, but either poorer or not improved estimates after experiencing prior-task failure. These effects of prior-task failure were not seen on estimation latencies as participants were faster during post-test than during pretest, most likely as a result of test-retest effects.

Effects of prior-task failure on strategy use

Better strategy use Older adults selected the better strategy more often than young adults (81.4% vs. 74.0%; F(1,116)=6.475, p=.01, MSe=505.0, η^2_p=.05). Moreover, the Condition × Testing interaction was significant (F(1,116)=20.597, p<.001, MSe=33.12, η^2_p=.15). Participants selected the better strategy more often during post-test than during pretest in the control condition (F(1,116)=8.012, p<.001, MSe= 38.42, η^2_p=.07), but less often in the failure condition F(1,116)=12.872, p< .001, MSe=38.42, η^2_p=.10), and this was the case in both young and older adults (Fs>7.361).

[1] For each variable (i.e., percentages of deviation, estimation latencies, and strategy use), differences between control and failure conditions during pretest were non-significant for both young and older adults (Fs<1.0).

As can be seen from Table 2, older adults decreased their use of the better strategy by 4.5% during post-test relative to pretest in the failure condition ($F(1,29)=6.636$, $p=.013$, $MSe=46.3$, $\eta^2_p=.24$), but used it equally often during pretest and post-test in the control condition ($F=1.591$, $p=.212$). Young adults increased their use of the better strategy by 3.0% during post-test relative to pretest in the control condition, $F(1,29)=10.461$, $p=.002$, $MSe=20.0$, $\eta^2_p=.23$, but decreased it by 3.0% in the failure condition, $F(1,29)=6.28$, $p=.011$, $MSe=20.0$, $\eta^2_p=.19$.

In other words, young and older adults used the better strategy either equally or more often during post-test than during pretest in the control condition, most likely as a result of test-retest effects, but less often in the failure condition, as a result of experiencing failure. Prior-task failure not only counteracted test-retest effect benefits but also led both young and older adults to select the better strategy less often.

Use of rounding down Participants selected the rounding-down strategy more often during post-test than during pre-test (62.3% vs. 59.3%; $F(1,116)=10.737$, $p=.001$, $MSe=81.3$, $\eta^2_p=.08$). This pre-/post-test difference was significant in both young ($F(1,58)=7.015$, $p=.010$, $MSe=121.0$, $\eta^2_p=.11$) and older adults ($F(1,58)=3.957$, $p=.045$, $MSe=41.5$, $\eta^2_p=.06$). Also, participants used the rounding-down strategy more often in the failure than in the control condition (62.8% vs. 55.4%; ($F(1,116)=7.667$, $p=.007$, $MSe=423.0$, $\eta^2_p=.06$). No other main or interaction effects were significant on mean percent use of the rounding-down strategy.

In summary, prior-task failure led both young and older adults to select the better strategy on each problem less often and to use the easier, rounding-down strategy more often. Rounding down is easier because it involves fewer processes: Participants do not need to increment decade digits; they calculate the product of decade digits that are displayed on the computer screen, and computations involve manipulating digits of smaller size. In all previous studies on computational estimation, rounding down was found to be executed more quickly by both young and older adults (e.g., Lemaire et al., 2004). Actually, participant-based correlations between mean percentages of better strategy use and mean percentages of rounding-strategy use during post-test were $rs=-.66$ and $-.74$ ($ps>.05$) in young and older adults, respectively, when tested under failure conditions (corresponding correlations were $rs=.22$, $.18$, $ps>.05$, under control conditions). These correlations differed between control and success conditions ($Zs>2.1$, $ps<.05$) but did not differ between young and older adults ($Zs<-.84$, $ps>.28$). Thus, following failure, participants used the better strategy on each problem less often and used the rounding-down strategy more often.

General discussion

We found evidence for effects of prior-task failure in both young and older adults. Participants' performance decreased after experiencing failure, and this decrease occurred above and beyond test-retest effects. These findings replicate previously reported effects of prior-task failure in young adults (Smith et al., 2006), and extend them to older adults. Also, by examining directly which strategy participants selected among two available strategies and their performance on each problem in a pretest–post-test design, our findings unambiguously establish that prior-task failure influences not only participants' performance but also their strategies. Such findings have important implications for further understanding and studying effects of prior-task failure specifically and of task transitions (prior-task success or failure) in general in young and older adults. We discuss these implications on when and how effects of prior-task failure occur, as well as on age-related differences and similarities in effects of prior-task failure.

When and how do effects of prior-task failure occur?

The present findings contribute to our understanding of effects of prior-task failure both empirically and theoretically. Empirically, our findings suggest that at least two features of task environment seem crucial for effects of prior-task failure to occur. First, participants' performance on the target task must not show floor effects. Here, participants selected the better strategy on more than 75% of problems during pretest. No decreased performance following prior-task failure can be found if participants perform too poorly on a given task.

Second, effects of prior-task failure may be less likely if this task is impossible to accomplish given time constraints. Recall that Geraci and Miller (2013) found no evidence for effects of prior-task failure. Most likely, given that participants in Geraci and Miller's experiment failed on prior task because they were given an unreasonably short time to accomplish it, participants may have felt that with enough time they could successfully accomplish this prior, sentence-scrambled task, leading them to not experience their lack of success as a true failure. Thus, participants in the failure condition likely discounted their failure on the prior task when accomplishing the target task. In other words, for effects of prior-task failure to occur, participants need to be able to succeed on the prior task, but actually fail. This is what happened here in our numerosity comparison task.

There are at least five aspects of effects of prior-task failure that may be further investigated in future research to understand their conditions of occurrence. One concerns whether prior and target tasks should be from the same or different domains. Here, our prior, numerosity comparison task was different from the target task, although it was from the same numerical processing domain. Smith et al. (2006) found that young adults' performance decreased on a target task when both prior and target tasks were the same (i.e., anagram problem solving). Previous findings from effects of prior-task success (e.g., Geraci et al., 2016; Lemaire & Brun, 2018; Lemaire et al., 2019) suggest that prior tasks should be cognitive if target tasks are from cognitive domains. Geraci et al. (2016) found no effects of prior task success when the target task was a cognitive task and the prior task was a motor task. Future research should determine which levels of overlap between prior and target tasks are required for effects of prior-task failure to occur.

Second, the present study did not assess how subjective rates of failure (or differences in participants' expected and actual performance) in the prior task influenced participants' performance and strategies in the target task. It would be interesting in future research to compare how subjective and objective (and differences between the two) rates of failure influence participants' performance and strategies in a target task. Objective rates refer to actual rates of failure (as assessed in this study), and subjective rates could be assessed by asking participants on how many trials they thought they failed (or succeeded in). Such an assessment would enable us to determine whether subjective experience of failure has the same effects as objective rates of failure, and whether differences between expected and actual performance in the prior task is one of the mechanisms underlying effects of prior-task failure.

Third, we ignore whether effects of prior-task failure are transient and occur only during the beginning of the target task or are more long-lasting and are found throughout the target task. Do they occur on only one target task or on several successive target tasks? For exploratory purposes, we compared participants' performance and percentages of better strategy use during the first 21 problems and the last 21 problems in the failure conditions. We found no differences between them. This suggests that effects of prior-task failure are not transient and last throughout target tasks. Nevertheless, with more problems in the target task, future studies may test more directly and strongly whether effects of prior-task success last throughout the target task or are seen only on initial trials. Also, future studies may investigate whether prior-task failure influences only the target task immediately following the prior task or whether they are lasting longer and impair performance on several successive target tasks.

Fourth, effects of prior-task failure were found here not only when participants failed but also when feedback was provided. Whether feedback is necessary for effects of prior-task failure to occur, and whether the contributions of feedback to effects of prior-task failure are the same or different in young and older adults are unknown. Some previous studies found effects of prior-task failure with no feedback (e.g., Smith et al., 2006), whereas other studies found no effects of prior-task failure with participants experiencing feedback (e.g., Geraci & Miller, 2013). Future studies should determine whether feedback is necessary for effects of prior-task failure to occur, and if they do, what type of feedback is necessary (on each trial and/or at the end of prior task), and whether experiencing feedback has the same contribution to effects of prior-task failure in young and older adults.

Fifth, future works may examine whether effects of prior-task failure differ (or not) for tasks or domains, like arithmetic tested here, that are less (or even not) age sensitive and for tasks and domains known to show much larger age-related decline (e.g., attention and episodic memory). It can be hypothesized that age differences in effects of prior-task failure are for tasks and domains in which older adults usually obtain poorer performance than young adults than in tasks or domains showing smaller age differences. Indeed, this could occur in domains like attention and memory if prior-task failure strengthens age-based stereotype threat, which is known to impair older adults' performance above and beyond aging effects (see Barber & Lui, 2020, for a review). Such a hypothesis may be tested in future works.

The theoretical contribution of the present findings concerns the mechanisms responsible for the effects of prior-task failure. After replicating the fact that prior-task failure decreases performance, the present findings have made one step forward in this direction by establishing that prior-task failure changes how participants accomplish target tasks. Experiencing failure led participants to select the better strategy less often on each problem and to use the easier, rounding strategy more often. In other words, prior-task failure changed the set of mechanisms used by participants to accomplish the target task. The next step is to determine how prior-task failure impacts strategy selection. Several types of mechanisms can be envisaged, ranging from cognitive to psychosocial mechanisms. Cognitive mechanisms include mechanisms that past research showed to be crucial in strategy selection, like executive control (e.g., Hinault et al., 2014, 2017; Lemaire & Lecaheur, 2010) or metacognition (Castel, Middlebrooks, & McGillivray, 2016; Geurten & Lemaire, 2019; Hertzog, 2016). For example, it is possible that prior-task failure leads participants to focus their attention away from processing task-relevant information like crucial stimulus features (e.g., size of unit digits in computational estimation problems). Psychosocial mechanisms include factors like increased stress and anxiety, decreased positive mood, or increased stereotype threat activation in older adults. For example, prior-task failure may increase participants' anxiety or failure-related concerns, which in turn interfere with processing target task. Speculatively, it could be envisaged that prior-task failure leads participants to be less confident in their cognitive performance, to increase failure-related concerns, and to lower their performance expectations. Such decreased self-confidence, increased failure-related concerns, or lowered performance expectations may lead participants to deploy fewer cognitive resources (e.g., less efficient executive control, smaller amount of attention, and/or less efficient monitoring of task performance) in the target task. This would result in lower performance by using the better strategy less often. Future studies may investigate candidate strategy selection mechanisms to determine how prior-task failure leads participants to obtain poorer performance and select better strategy less often. Above and beyond deciphering mechanisms underlying how prior-task failure leads to poorer performance, this would enable us to know how cognitive and non-cognitive factors contribute to specific effects of prior-task failure as well as to more general effects associated with situational parameters (e.g., task transition, speed-accuracy pressures, emotions, stereotype threat) that are known to either modulate age-related differences in human cognition or not.

Aging and effects of prior-task failure

Surprisingly, our data showed comparable effects of prior-task failure in young and older adults, and that these effects occurred via the same mechanisms in both age groups. For several reasons, it could be expected that experiencing failure would have stronger detrimental effects in older adults (e.g., older adults' confidence may be more sensitive to failure; failure may have strengthened an age-based stereotype threat known to lead older adults to underperform). Including arithmetic fluency scores as covariates revealed that our older adults' higher arithmetic fluency did not buffer potential age-related differences in effects of prior-task failure.

Our age equivalence findings do not mean that there may be no age differences in effects of prior-task failure. Testing different levels of failure (with some participants being tested in conditions where careful selection of items leads them to fail on 30%, 60%, or 90% of trials), all else being equal, may reveal age differences that did not occur in the present context. Also, before too prematurely concluding age invariance in effects of prior-task failure, it is important to test other cognitive domains or tasks. Age-related differences in effects of prior-task failure might occur in other, more age-sensitive domains (e.g., episodic memory). This would enable us to determine whether age differences and similarities in effects of prior-task failure interact with cognitive domains or tasks.

Note that our findings of age invariance in effects of prior-task failure contrasts with effects of prior-task success that previous studies found to be stronger in older than in young adults (Geraci & Miller, 2013; Geraci et al., 2016; Lemaire & Brun, 2018; Lemaire et al., 2019). Prior-task success leads to increased performance whereas prior-task failure impairs performance on a target task. This suggests that older adults may be more sensitive than young adults to task transitions (e.g., prior-task success) that facilitate their performance, and that both age groups may be equally sensitive to task transitions (e.g., prior-task failure) that impair cognitive performance. Such a general conclusion awaits further evidence from different cognitive domains and tasks as well as from different levels of prior-task failures.

Acknowledgements The author would like to thank Mikael Flatot for his help in analyzing the data.

Open practices statement The raw data of this study can be found at https://osf.io/u9awp/. This experiment was not pre-registered.

Funding This research was supported by grants from the Agence Nationale de la Recherche (Grant # ANR-17-CE28-0003; ANR # 20-ASTC-0031).

References

Abrams, D., Crisp, R. J., Marques, S., Fagg, E., Bedford, L., & Provias, D. (2008). Threat inoculation: Experienced and imagined intergenerational contact prevents stereotype threat effects on older people's math performance. *Psychology and Aging*, *23*(4), 934-939. https://doi.org/10.1037/a0014293.

Barber, S. J. (2017). An Examination of Age-Based Stereotype Threat About Cognitive Decline: Implications for Stereotype-Threat Research and Theory Development. *Perspectives on Psychological Science*, *12*(1), 62-90. https://doi.org/10.1177/1745691616656345.

Barber, S.J., & Lui, K. (2020). Stereotype threat and the cognitive performance of older adults. In A. Thomas & A. Gutchess (Eds.), *Handbook of Cognitive Aging: A Life Course Perspective,* Cambridge University Press. (pp. 400-416).

Beilock, S. L. (2008). Math Performance in Stressful Situations. *Current Directions in Psychological Science*, *17*(5), 339-343. https://doi.org/10.1111/j.1467-8721.2008.00602.x

Beilock, S. L., & DeCaro, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 983-998. https://doi.org/10.1037/0278-7393.33.6.983.

Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, *36*(3), 604-616. https://doi.org/10.3758/MC.36.3.604.

Campbell, J. I. D., & Austin, S. (2002). Effects of response time deadlines on adults' strategy choices for simple addition. *Memory & Cognition*, *30*(6), 988-994. https://doi.org/10.3758/BF03195782.

Castel A.D., Middlebrooks C.D., McGillivray S. (2016). Monitoring memory in old age: impaired, spared, and aware. In: Dunlosky J, Tauber S, editors. The Oxford Handbook of Metamemory. Oxford: Oxford University Press (pp. 519–36). https://doi.org/10.1093/oxfordhb/9780199336746.013.3.

Cohen Kadosh, R. & Dowker, A. (2015). *The Oxford handbook of numerical cognition*. Oxford, UK: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642342.001.0001.

Deltour, J. J. (1993). *Echelle de Vocabulaire Mill Hill de J. C. Raven: Adaptation française et normes comparées du mill hill et du standard progressive matrices (PM38). Manuel et annexes.* (L'Application des Techniques Modernes SPRL). Braine le Château, Belgium.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191. https://doi.org/10.3758/BF03193146

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state. *Journal of Psychiatric Research*, *12*(3), 189-198. https://doi.org/10.1016/0022-3956(75)90026-6

French J.W., Price L.A., Akstrom R.B. (1963). *Manual for kit of reference tests for cognitive factors*. Educational Testing Services.

Geraci, L., Hughes, M. L., Miller, T. M., & De Forrest, R. L. (2016). The Effect of Prior Task Success on Older Adults' Memory

Performance: Examining the Influence of Different Types of Task Success. *Experimental Aging Research*, *42*(4), 365-381. https://doi.org/10.1080/0361073X.2016.1191860.

Geraci, L., & Miller, T. M. (2013). Improving older adults' memory performance using prior task success. *Psychology and Aging*, *28*(2), 340-345. https://doi.org/10.1037/a0030332

Geurten M, Lemaire P. (2019) Metacognition for strategy selection during arithmetic problem-solving in young and older adults. Aging, Neuropsychology, and Cognition, 25(3), 424-446, https://doi.org/10.1080/13825585.2018.1464114.

Gilmore, C., Göbel, S. M., & Inglis, M. (2018). *An introduction to mathematical cognition*. London; New York: Routledge, Taylor & Francis Group. https://doi.org/10.4324/9781315684758.

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proceedings of the National Academy of Sciences*, *109*(28), 11116-11120. https://doi.org/10.1073/pnas.1200196109.

Haslam, C., Morton, T. A., Haslam, S. A., Varnes, L., Graham, R., & Gamaz, L. (2012). "When the age is in, the wit is out": Age-related self-categorization and deficit expectations reduce performance on clinical tests used in dementia assessment. *Psychology and Aging*, *27*(3), 778-784. https://doi.org/10.1037/a0027754.

Hertzog C. (2016). Aging and metacognitive control. In: Dunlosky J, Tauber S, (Eds). *The Oxford Handbook of Metamemory*. Oxford: Oxford University Press (pp. 537–558).

Hinault T, Dufau S, Lemaire P. (2014) Sequential modulations of poorer-strategy effects during strategy execution: An event-related potential study in arithmetic. *Brain & Cognition*, *91,* 123–30. https://doi.org/10.1016/j.brainres.2015.10.057

Hinault, T., Badier, J.-M., Baillet, S., & Lemaire, P. (2017). The Sources of Sequential Modulations of Control Processes in Arithmetic Strategies : A Magnetoencephalography Study. *Journal of Cognitive Neuroscience*, *29*(6), 1033-1043. https://doi.org/10.1162/jocn_a_01102.

Lamont, R. A., Swift, H. J., & Abrams, D. (2015). A review and meta-analysis of age-based stereotype threat: Negative stereotypes, not facts, do the damage. *Psychology and Aging*, *30*(1), 180-193. https://doi.org/10.1037/a0038586.

Lemaire, P. (2010). Cognitive Strategy Variations During Aging. *Current Directions in Psychological Science*, *19*(6), 363-369. https://doi.org/10.1177/0963721410390354

Lemaire, P. (2016). *Cognitive aging: The role of strategies*. London; New York: Routledge, Taylor & Francis Group. https://doi.org/10.4324/9781315650999

Lemaire, P., Arnaud, L., & Lecacheur, M. (2004). Adults' Age-Related Differences in Adaptivity of Strategy Choices: Evidence From Computational Estimation. *Psychology and Aging*, *19*(3), 467-481. https://doi.org/10.1037/0882-7974.19.3.467

Lemaire, P., & Brun, F. (2016). Age-related differences in sequential modulations of problem-size and rule-violation effects during arithmetic problem verification tasks. *Memory & Cognition*, *44*(3), 444-453. https://doi.org/10.3758/s13421-015-0566-7

Lemaire, P., & Brun, F. (2018). Effects of Prior-Task Success on Young and Older Adults' Cognitive Performance an Evaluation of the Strategy Hypothesis. *Journal of Cognition*, *1*(1), 14. https://doi.org/10.5334/joc.17

Lemaire, P., Gouraud, J., & Nicolas, P. (2019). When Older Adults Outperform Young Adults: Effects of Prior-Task Success in Arithmetic. *Gerontology*, 1-10. https://doi.org/10.1159/000500673

Lemaire P, Lecacheur M. (2010). Strategy switch costs in arithmetic problem solving. *Memory & Cognition, 38(3),*322–32. https://doi.org/10.3758/mc.38.3.322

Lemaire, P., & Leclère, M. (2014). Strategy repetition in young and older adults : A study in arithmetic. *Developmental Psychology*, *50*(2), 460-468. https://doi.org/10.1037/a0033527

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), 364-382. https://doi.org/10.3758/BF03211420

Mauro, D. G., LeFevre, J.-A., & Morris, J. (2003). Effects of problem format on division and manipulation performance: Division facts are mediated via multiplication-based representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 163-170. https://doi.org/10.1037/0278-7393.29.2.163

Nicolas, P., Lemaire, P., & Régner, I. (2020). When and how stereotype threat influences older adults' arithmetic performance: Insight from a strategy approach. *Journal of Experimental Psychology: General*. *149*(2), 343–367. https://doi.org/10.1037/xge0000647

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales: The Mill Hill Vocabulary Scale* (Harcourt Assessment). San Antonio.

Roquet, A., & Lemaire, P. (2019). Strategy variability in numerosity comparison task: A study in young and older adults. *Open Psychology*, *1*(1), 152-167. https://doi.org/10.1515/psych-2018-0011

Schneider, D. W., & Anderson, J. R. (2010). Asymmetric Switch Costs as Sequential Difficulty Effects. *Quarterly Journal of Experimental Psychology*, *63*(10), 1873-1894. https://doi.org/10.1080/17470211003624010

Smith, S. A., Kass, S. J., Rotunda, R. J., & Schneider, S. K. (2006). If at first you don't succeed: Effects of failure on general and task-specific self-efficacy and performance. *North American Journal of Psychology*, *8*(1), 171-182.

Strickland-Hughes, C. M., West, R. L., Smith, K. A., & Ebner, N. C. (2017). False feedback and beliefs influence name recall in younger and older adults. *Memory*, *25*(8), 1072-1088. https://doi.org/10.1080/09658211.2016.1260746

Uittenhove, K., & Lemaire, P. (2013a). Strategy sequential difficulty effects in Alzheimer patients: A study in arithmetic. *Journal of Clinical and Experimental Neuropsychology*, *35*(1), 83-89. https://doi.org/10.1080/13803395.2012.753036

Uittenhove, K., & Lemaire, P. (2013b). Strategy sequential difficulty effects vary with working-memory and response–stimulus-intervals: A study in arithmetic. *Acta Psychologica*, *143*(1), 113-118. https://doi.org/10.1016/j.actpsy.2013.02.007.

Uittenhove, K., & Lemaire, P. (2015). Numerical cognition during cognitive aging. In R. Cohen Kadosh & A. Dowker (Eds.), *Oxford Handbook of Numerical Cognition*. Oxford. *109*(28), 345-364. https://doi.org/10.1093/oxfordhb/9780199642342.013.045

Uittenhove, K., & Lemaire, P. (2018). Performance Control in Numerical Cognition: Insights from Strategic Variations in Arithmetic during the lifespan. In Henik, A., & Fias, W. (Eds.), *Heterogeneity of Function in Numerical Cognition*, Elsevier: Academic Press (pp. 127—145). https://doi.org/10.1016/b978-0-12-811529-9.00007-8.

West, R. L., Dark-Freudeman, A., & Bagwell, D. K. (2009). Goals-feedback conditions and episodic memory: Mechanisms for memory gains in older and younger adults. *Memory*, *17*(2), 233-244. https://doi.org/10.1080/09658210802236407.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.