When and how stereotype threat influences older adults arithmetic performance? Insight from a strategy approach

Poshita Nicolas, Patrick Lemaire, Isabelle Regner

When and how stereotype threat influences older adults' arithmetic performance? Insight from a strategy approach.

Poshita Nicolas, Patrick Lemaire, & Isabelle Régner

Aix Marseille Univ, LPC, CNRS, Marseille, France

Corresponding author:
Patrick Lemaire
CNRS & Aix-Marseille Université
3 Place Victor Hugo
13331 Marseille France
Email: patrick.lemaire@univ-amu.fr

Acknowledgments

This research was supported in part by the CNRS (French NSF), a grant from the Agence Nationale de la Recherche (Grant # ANR-17-CE28-0003-01-01) to PL, and a grant from the Agence Nationale de la Recherche (GRANT # ANR-16-CE36-0005-01) to IR. Correspondence about this paper should be directed to Patrick Lemaire, Aix-Marseille University & CNRS, 3 Place Victor Hugo, Case D, 13331 Marseille, France (email: Patrick.Lemaire@univ-amu.fr).

Word count = 17124

Abstract

In three experiments, we investigated how age-related differences in cognitive performance are exacerbated by age-based stereotype threat. We adopted a strategy approach and investigated a domain, namely arithmetic, where age-related differences in participants' performance are either non-existent or very small and where effects of age-based stereotype threat have never been investigated. In two types of tasks (problem verification in Expt. 1 and computational estimation in Expts. 2 & 3), we found that age-based stereotype threat led older adults to obtain poorer performance, to adopt less systematically and less often the better strategy on each problem, to repeat the same strategy across trials even when it was inappropriate, and to execute available strategies more poorly. We also found that poorer strategy use mediated threat effects and that individual differences in processing resources moderated individuals' sensitivity to effects of stereotype threat. Our results establish that age-based stereotype threat effects are independent of the cognitive domain or tasks in which they are studied and of pre-experimental differences in young and older adults' performance. They deepen our knowledge of the mechanisms underlying age-based, and other, stereotype threat effects. They also document how domain-general and domain-specific processing resources moderate individual differences in age-based stereotype threat effects. Our findings have important implications to improve our understanding of how and when age-based (and other) stereotype threat effects occur, and, more generally, how psychosocial factors modulate age-related changes in human cognition.

Keywords: Cognitive Aging, Stereotype Threat, Strategies, Arithmetic.

Introduction

Negative aging stereotypes (e.g., older adults viewed as being slow, incompetent, forgetful, and senile) are widely held in our society (e.g., Kite & Johnson, 1988; Stein, Blanchard-Fields, & Hertzog, 2002; see Hummert, 1999; Kite, Stockdale, Whitley, & Johnson, 2005, for reviews). These stereotypes make older adults vulnerable to age-based discrimination and put them at the risk of experiencing stereotype threat - a predicament that occurs when people encounter situations in which they feel they might confirm the negative stereotype about their group (Steele & Aronson, 1995). It is now well established that age-based stereotype threat (ABST) can lead to significant decrements in older adults' cognitive performance (see Barber, 2017; Barber & Mather, 2014; Chasteen, Kang, & Remedios, 2012; Lamont, Swift, & Abrams, 2015, for reviews). However, some important issues remain. First, mechanisms underlying ABST are still at a speculative stage. Second, unknown is whether some cognitive resources might help older adults resist (or be less vulnerable) to ABST effects. Third, whether ABST is powerful enough to affect older adults' cognitive performance in domains undergoing no (or much less) age-related declines is an open question. We address these issues in the present research and thus provide new insights to further understand ABST effects.

Effects such as ABST are important to study when we try to understand age-cognition relations, as they suggest that psychosocial factors have a crucial role on age-related deficits on cognition. These psychosocial factors influence older adults' cognitive performance above and beyond general cognitive factors already established, like slower processing rates (Salthouse, 1996), decreased working memory (Foos, 1989), declines in attentional resources (Craik & Byrd, 1982), or sensory deficits (Baltes & Lindenberger, 1997). Here, in three experiments, we investigated strategic aspects underlying ABST effects and tested potential cognitive moderators. We carried out these experiments in the domain of arithmetic, where young and older adults usually show comparable baseline performance. Before outlining the logic of the present project, we review previous relevant findings on ABST effects and on aging and strategic variations.

Detrimental impact of ABST has been robustly found on cognitive tests typically used in lab experiments (e.g., Hess, Auman, Colcombe, & Rahhal, 2003; Mazerolle, Régner, Morisset, Rigalleau, & Huguet, 2012) as well as on short cognitive tests used in clinical settings for dementia screening (e.g., Barber, Mather, & Gatz, 2015; Mazerolle et al., 2017). Although most ABST effects have been investigated in the domain of memory (see Armstrong, Gallant, Patel, & Wong, 2017; Barber, 2017; Barber & Mather, 2014; Lamont et al., 2015, for reviews), they have also been found in other domains such as physical abilities, skill acquisition, and driving (Abrams, Eller, & Bryant, 2006; Gaillard, Desmette, & Keller, 2011; Haslam et al., 2012; Joanisse, Gagnon, & Voloaca, 2013; Lambert et al., 2016; Swift, Lamont, & Abrams, 2012). ABST detrimental effects are readily observable and fairly easy to produce with instructional manipulations (Haslam et al., 2012), such as simply emphasizing the memory component of the test (Desrichard & Kopetz, 2005; Kang & Chasteen, 2009; Rahhal, Hasher, & Colcombe, 2001), highlighting differences in performance between young and older adults (Hess et al., 2003; Hess & Hinson, 2006), or implicitly activating negative aging stereotypes (Levy, 1996; Stein et al., 2002). Likewise, older adults' performance can be restored under some conditions, like when instructions de-emphasize the memory component of the test (e.g., Desrichard & Kopetz, 2005; Rahhal et al., 2001) or when the task is presented as usually showing no-age differences (e.g., Brubaker & Naveh-Benjamin, 2018; Hess et al., 2003; Mazerolle et al. 2012; Mazerolle, Régner, Rigalleau, & Huguet, 2015).

Previous studies have investigated ABST effects in domains where pronounced age-related differences have been already established, like memory (Old & Naveh-Benjamin, 2008), motor behaviour (Voelcker-Rehage, 2008), skill acquisition (Head, Raz, Gunning-Dixon, Williamson, & Acker, 2002), and driving skills (Chaparro & Alton, 2000). However, an interesting issue is whether ABST is powerful enough to also occur in domains where older adults are as proficient as young adults or where age-related differences are smallest. Some research conducted on gender stereotypes in STEM fields found stereotype threat effects (i.e., decreased math performance in women relative to men) in tasks showing no baseline differences in men's and women's performance (Pavlova, Wecker, Krombholz, & Sokolov, 2010; Pavlova, Weber, Simoes, & Sokolov, 2014).

It is important to determine whether ABST effects also occur in cognitive domains where age-related differences are inexistent or smaller for several reasons. It would enable us to know if ABST effects exist independently of older adults' baseline performance like other stereotype threat effects or rather function with specificities inherent to aging processes. It would enable to compare magnitudes of ABST effects in domains where age-related differences in baseline performance exist and in domains where they do not exist (or are much smaller). It would also contribute to determine whether older adults' high proficiency in one domain gives them resources to resist stereotype threat effects. Finally, it would help to better estimate how stereotype threat contributes to age-related differences in cognitive performance. We addressed these issues here by testing ABST in arithmetic, a domain where age-related differences are either non-existent or much smaller (see Duverne & Lemaire, 2005; Uittenhove & Lemaire, 2015, 2018, for reviews).

The mathematical domain has been the focus of many gender stereotype threat research in STEM fields (e.g., Beilock, Rydell, & McConnell, 2007; Beilock, 2008; Huguet & Régner, 2007, 2009; Schmader, 2002; Schmader & Johns, 2003; Spencer, Logel, & Davies, 2016; Spencer, Steele, & Quinn, 1999; Quinn & Spencer, 2001; Smeding, Dumas, Loose, & Régner, 2013), while memory has been the main focus of ABST research. To our knowledge, only one study investigated ABST in the math domain (Abrams et al., 2008), and showed that older adults' math performance was impaired when differences in performance between young and older adults were highlighted. However, there were two limitations in this study. First, participants had to solve only 15 problems, and the type and list of problems were not provided. So, it is impossible to know what type of arithmetic processes was influenced by ABST. Second and more importantly, individuals' arithmetic fluency was not assessed independently of the target math test. It is thus impossible to know whether group differences were the result of ABST manipulations or of group differences in baseline arithmetic performance. The present study tested many more arithmetic problems, the solution process of which has been investigated in previous research examining aging effects on arithmetic performance. We also assessed individuals' arithmetic fluency with a test that was independent of the experimental arithmetic tasks and controlled that ABST effects were not contaminated by group differences in arithmetic fluency.

An additional reason to examine stereotype threat effects in arithmetic concerns the mechanisms underlying ABST effects, as findings from arithmetic shed important light on both domain-general and domain-specific cognitive processes as well as on age-related changes in these processes (see Kadosh & Dowker, 2015; Gilmore, Göbel, & Inglis, 2018, for overviews). Currently, there is no consensus on how stereotype threat impairs older adults' cognitive performance. Research with young adults in the stereotype threat literature has found anxiety (e.g., Steele, 1997; Steele, Spencer, & Aronson, 2002; Osborne, 2001), depletion of executive control resources (e.g., Schmader & John, 2003; Schmader, Johns, & Forbes, 2008), and an imbalance in regulatory fit (Barber & Mather, 2013a, 2013b; Grimm, Markman, Maddox, & Baldwin, 2009; Seibt & Förster, 2004) to mediate stereotype threat effects. However, support for these mechanisms to explain threat effects in older adults are either inconsistent or insufficient (see Barber, 2017; Pennington, Heim, Levy, & Larkin, 2016, for reviews). Mixed findings have also been reported for other mediating variables in the ABST literature, like perceived stereotype threat (see Chasteen, Bhattacharyya, Horhota, Tam, & Hasher, 2005, for an example of significant effects; and Barber et al., 2015, for an example of null effects) and lower performance expectation (see Desrichard & Kopetz, 2005 for an example of effects; and Chasteen et al., 2005, for example of null effects).

A recently proposed promising perspective on stereotype threat effects in general and ABST effects in particular is the strategy perspective. This strategy perspective is based on previous findings that people use several strategies to accomplish cognitive tasks (Siegler, 2007), and that condition-related and age-related differences in participants' performance are systematically associated with strategic variations (see Lemaire, 2016, for an overview). A strategy is "a procedure or set of procedures for achieving a higher level goal or task" (Lemaire & Reder, 1999, p. 365). According to the strategy perspective, ABST effects occur via impairing older adults' ability to select the best strategy on each item and/or to execute strategies efficiently. One study found some indirect evidence suggesting that threat reduced the use of more efficient memory strategies (Hess et al., 2003). Also, only one study to date has directly tested this strategy account of ABST effects. Lemaire, Brun, and Régner (2018) asked older adults to encode lists of pairs of words using either a repetition strategy or an imagery strategy in threat and control conditions. In one experiment, participants were free to choose between the repetition or the imagery strategy to encode each pair of words, and indicated which strategy they chose after each pair of words. In another experiment, the authors asked one group of older adults to encode all pairs of words with a repetition strategy and another group of older adults to encode all pairs of words with the imagery strategy. Besides threat effects on performance in a cued-recall task, Lemaire et al. found that ABST disrupted both the selection and the execution of the most efficient, but also the most demanding, imagery strategy in older adults. One of the limitations of this study was that ABST effects on strategy execution were assessed on recall performance, that is indirectly assuming that better recall resulted from better execution of encoding strategy. Although this is plausible, the present studies tested strategy execution directly by examining how fast and accurate participants are when solving arithmetic problems with a given strategy. Additionally, examining the strategy hypothesis here in arithmetic was a stronger test of the role of strategies in ABST effects, as strategy differences between control and ABST groups in the Lemaire et al.'s study may be specific to the memory domain where older adults have poorer capacity than young adults.

We conducted three studies to test the strategy hypothesis of ABST, one in which strategies were investigated indirectly (Expt. 1), the other two in which they were investigated directly (Expts. 2 & 3). It is important to investigate strategies both directly and indirectly for several reasons. First, in most cognitive tasks (like the arithmetic problem verification task used here; Expt. 1), strategies that people use cannot be directly observed (i.e., with external behavioural evidence of strategy use independently of participants' performance) and thus conclusions about strategy use are inferred from the patterns of speed and accuracy that arise as a function of the factors that define the stimulus set. Second, some tasks (like the computational estimation task used here; Expts. 2 & 3) enable to investigate strategy use and strategy execution directly and independently, given that they yield behavioral external evidence that increase our confidence in how participants accomplish cognitive tasks and in the fact that participants use different strategies and/or execute them differently in different conditions. Finally, in some tasks like the computational estimation task used here, it is also possible to instruct participants which strategy to use on each item and control that they comply with strategy instructions.

The final important goal of our study was to investigate whether stereotype threat can be manipulated within the same participant by adopting a pretest-posttest design. A recent study used a within-participants design and demonstrated ABST effects on memory performance in the same individuals tested under nonthreat and threat conditions (Brubaker and Naveh-Benjamin, 2018). We wanted to replicate such findings and extend them to arithmetic, as such findings importantly speak to the very nature of stereotype threat, namely that, as assumed by

Steele (1997), that situational threat can be induced in important situations. Our additional objectives in adopting such a pretest-posttest design were to document changes in strategic variations within the same individuals from pretest to posttest condition in the threat and control groups and to determine whether all older adults experience negative consequences of ABST to similar extents or whether some older individuals suffer more than others from ABST. Previous research suggests important individual differences in ABST effects. Older adults who are most likely to show performance impairment when faced with aging stereotypes are those who identify strongly to the stereotyped group and the stereotyped domain (Hess et al., 2003), are more highly educated (Hess, Hinson, & Hodges, 2009), show high levels of stigma consciousness or perceived stereotype threat (Hess et al., 2009; Kang & Chasteen, 2009), belong to the young-old (60-70 years) relative to old-old (71-85 years) age group (Hess & Hinson, 2006; Hess et al., 2009), experience or imagine poorer and/or less numerous intergenerational interactions (Abrams et al., 2006; Abrams et al., 2008). Here, we examined whether some older individuals are much less negatively affected than others by ABST

In summary, we pursued three goals in the present research. First, we wanted to determine whether older adults are sensitive to ABST when they are tested in cognitive domains, like arithmetic, where young and older adults' baseline performance are equally good, in contrast to domains like memory that undergoes important age-related declines. Second, to further our knowledge of mechanisms underlying ABST effects, we asked whether older adults perform poorly under stereotype threat because threat leads them to use different sets of mechanisms (strategies) and/or use the same set of mechanisms but execute them poorly. Third, we investigated these changes in strategy use and strategy execution in a within-participants design. Moreover, we investigated individual differences in ABST.

Experiment 1

Experiment 1 aimed at determining whether older adults are sensitive to ABST when tested in a domain where they are as competent as young adults. We asked young and older adults to accomplish arithmetic problem verification tasks. In these tasks, participants are given arithmetic equations and are asked to say as quickly and as accurately as possible whether equations are true or false (e.g., 4 x 17 = 68, True or False?). Because previous studies (Krueger, 1986; Kruegar & Hallford, 1984) have illustrated that semantic features of arithmetic problems (a) are used by participants to select strategies that improve their performance, and (b) determine the way participants execute strategies, we manipulated our arithmetic problems to use two semantic features, parity (in false problems) and problem-size (in true problems) to investigate threat effects on strategy use and strategy execution. Strategy execution was examined by determining problem-size effect in true problems and strategy use was examined by determining parity violation effect in false problems.

~~Previous works revealed that arithmetic problem verification tasks can be fruitfully used to document strategy use and strategy execution.~~ Strategy execution can be examined by determining problem-size effect (i.e., ~~comparing~~ performance difference on two types of true problems small and large problems; see Zbrodoff & Logan, 2005, for a review) ~~that participants are known to solve with the same strategy~~. Previous studies reported that participants use the same ~~an~~ exact-calculation strategy for verifying ~~these~~ both ~~true~~ small and large problems and are faster on small than on large problems (i~~.e., a robust effect in arithmetic known as problem size effect; see Zbrodoff & Logan, 2005, for a review~~), further implying that the exact-calculation strategy is executed differently on small and large problems. The exact-calculation strategy involves encoding the operands and proposed answer, searching for the correct answer (via either calculation or retrieval), comparing the correct and proposed answers, making a true/false decision, and then responding by pressing the appropriate response key. Because answers to larger problems are harder to find, participants take more time to verify true large problems.

Strategy use in arithmetic problem verification tasks can be examined by determining parity violation effect (i.e., difference in performance on two types of false problems, match and mismatch problems). ~~comparing participants' performance on different types of false problems. For example,~~ Previous studies found that participants are faster on false problems if the odd-even status of the proposed and correct answers mismatches (e.g., 13x7=92) than if it matches (e.g., 13x7=93). This so-called parity violation effect has been found in both young and older adults (Anders, Hinault, & Lemaire, 2018; Hinault & Lemaire, 2017; Hinault, Dufau, & Lemaire, 2015a; Hinault, Tiberghien, & Lemaire, 2015b; Krueger & Hallford, 1984; Krueger, 1986; Lemaire & Reder, 1999; Lemaire & Fayol, 1995; Masse & Lemaire, 2001). It has been accounted for as a result of participants using two different strategies on each type of parity-match and parity-mismatch problems. Participants use a time-consuming exhaustive verification strategy like the one used for verifying true problems on parity-match problems and a fast, plausibility checking strategy on parity-mismatch problems. With this plausibility checking strategy, participants quickly decide that the equation is false because they quickly detect that the odd-even status of the proposed and correct answers is different, which is enough information to make a correct true/false decision. In using the fast plausibility-checking strategy, participants do not complete the search of the correct answer and use other semantic features (e.g., parity of operands and proposed answers) to quickly reject the parity-mismatch problems.

In the present experiment, above and beyond ABST effects on older adults' performance, we tested the hypothesis that ABST will impair older adults' strategy use and strategy execution. Changes in older adults' strategy use and strategy execution were tested by determining how magnitudes of parity violation effects (strategy use) and of problem-size effects (strategy execution) vary as a function of ABST. The hypothesis that ABST influences strategy execution predicts increased problem-size effects in older adults when verifying true problems under stereotype threat. Moreover, the hypothesis that ABST influences strategy use predicts decreased parity effects in older adults when verifying false problems under threat.

Method

Participants.

Eighty young and 80 older adults participated in Experiment 1. Older adults were volunteers from distinct French metropolitan areas, and young adults were undergraduate students at Aix-Marseille University who received course credit for their participation. Prior to the experiment, all older adults accomplished the Mini Mental-State Examination (MMSE, Folstein, Folstein & McHugh, 1975), and none obtained an MMSE score < 27. Participants' characteristics are summarized in Table 1. The target sample size was determined using an a-priori power analysis (G*Power; Faul, Erdfelder, Lang, & Buchner 2007). In their meta-analysis of ABST, Lamont et al. (2015) found an effect size of $d = .52$ when using stereotype-based manipulations as we did here. Using this reported effect size (corresponding to $f = .25$), the error rate set to 0.05, and the power set to 0.80, the power analysis indicated that a sample of 128 participants (32 per conditions) would be sufficient to detect the critical 2(Age: young, older adults) x 2(Group: threat, control) between-participants interaction (i.e., ABST effects on arithmetic performance). This sample size was also large enough to detect ABST impact on either problem-size effects or parity effects, which both involve probing for a between-within interaction (that typically requires a smaller sample size). To achieve greater than 80% power, we decided to recruit 40 participants per group, resulting in a final sample of 160 participants.

Table 1. Participants' Characteristics in Threat and Control Groups

	Young Adults			Older Adults		
Variables	Threat *M(SD)*	Control *M(SD)*	*p*	Threat *M(SD)*	Control *M(SD)*	*p*

N (Females)	40 (30)	40 (28)	--	40 (28)	40 (26)	--
Age (Years)	21 (3.4)	20.2 (2)	0.64	70 (5.3)	69 (3.5)	0.21
Education (Years)	14 (1.7)	13.5 (1.1)	0.17	14.6 (1.5)	14.3 (1.9)	0.51
MMSE	--	--	--	29.5 (0.6)	29.2 (0.8)	0.07
Arithmetic Fluency Test	30.5 (10.5)	25.7 (12.7)	0.22	72.5 (11.9)	86.9 (26.8)	<.001
MHVS	22 (4.6)	21.3 (3.7)	0.38	27.8 (4.3)	27.7 (3.4)	0.59

Stimuli.

The stimuli were 288 multiplication problems presented in a standard form (i.e., a x b = c), with the operands a and b being either single or double digit numbers. Single-digit operands ranged from 3 to 9, whereas double-digit operands ranged from 12 to 97. Two types of multiplication problems were presented: True or False problems. The basic set of equations consisted of 144 unique multipliers. Forty-eight unique problems were selected, two-even, one-even, or zero-even (two-odd) operands. All true problems had the same operands as false problems and differed only in the value given as the proposed product. This value was the correct product of the two operands for true problems. Half the problems were large problems and half were small problems. Correct products ranged from 48 to 204 for small problems (*mean*=139) and from 207 to 441 for large problems (*mean*=301).

Two types of false problems were tested: (a) *Parity-Match* (or match) problems involved false answers with odd-even status that were the same as those of the correct products (e.g., 19 x 7 = 131), (b) *Parity-Mismatch* (or mismatch) problems involved false answers with odd-even status that were different from those of correct products (e.g., 19 x 7 = 132). All false problems were created by varying splits (i.e., differences) between correct and proposed products. Incorrect answers were off by ±1, ±2, or ±3 from correct answers. For each group of zero-even, one-even, and two-even operand problems, there were 48 unique problems, 24 *Match* and 24 *Mismatch* problems, six each from splits of ±1 and ±3 for mismatch problems, and 12 each from splits of ±2 for match problems.

Based on previous findings in arithmetic (see Kadosh & Dowker, 2015; Gilmore et al., 2018, for overviews), we controlled the following factors: (a) split size (i.e., the sum of positive and negative splits equaled zero for match and mismatch problems), (b) no double-digit operand had zero or five as unit digit (e.g. 30, 45), (c) no double-digit operand had the same unit and decade digits (e.g. 22, 44), (d) the size and side of operands were controlled, such that all problems had both a single digit and a double digit operand and that half the problems had the double-digit in the left position (e.g., 17 x 6 = 102), and half in the right position (e.g., 4 x 17 = 68), (e) all problems with only one even operand had half of their even multiplicand in the right position (e.g., 17 x 4 = 68) and the other half in the left position (e.g., 6 x 13 = 78), (f) none of the problems included zero, one, or five as a single-digit operand, (g) two was never used as single-digit operand to avoid discrepancy in the number of odd (3, 7, 9) and even digits (4, 6, 8), and finally (h) the multiplicand and the multiplier were never the same (e.g., 8 x 8 = 64 or 12 x 12 = 144).

Procedure.

Participants were tested individually in one session that lasted approximately 90 minutes. Half the participants were randomly assigned to the Threat group and half to the Control group.

Threat Group. Our stereotype threat manipulation was adapted from the previous studies (e.g., Mazerolle et al. 2012, 2015; Inzlicht & Ben-Zeev, 2000; Lemaire et al., 2018; Schmader & Johns, 2003). The experimental session was introduced as a study of memory and attention abilities. The experimenter was presented as a psychologist and expert in cognitive aging. Then, participants were told that they were going to perform a test relying heavily on their memory and attention abilities. The test included an arithmetic verification task which required them to verify, as quickly and accurately as possible, if the proposed multiplication equations (e.g., 17 x 6 = 102) were true or false. Participants were specifically told that they were evaluated for their speed and accuracy. In addition, participants were cued that they could only successfully complete the test if they relied adequately on their memory and attention abilities. Also, an objective of the study was presented so as to investigate any differences between young and older adults' performance on the test. Participants were therefore informed of the presence of a comparison group. After confirming that participants had thoroughly understood the objectives and demands of the present study, they were prompted to indicate their age and started the test on a laptop. The experimental session included two parts with a brief rest period of 5-10 minutes in-between each part. Before the participants resumed the task, the experimenter reminded them of (a) the objective of the study, (b) that they were still being evaluated for their speed and accuracy, (c) that to successfully accomplish the test they were to rely strongly on their memory and attention abilities, and (d) that older adults were compared to young adults.

Control Group. A procedure similar to that for the threat group was adopted for the control group, with the following exceptions. The session was introduced as a study on arithmetic. The experimenter was presented as a student. The task was not presented as a test. Instead, participants were merely informed that they were to perform an arithmetic problem verification task (e.g., 17 x 6 = 102) as quickly and accurately as they can. The speed component of the task was not emphasized. There was no mention of memory and attention abilities. None of the participants were prompted to indicate their age prior to the task. Age was recorded during participant recruitment. Participants were specifically told that the objective of the study was to understand how people in general solve and verify multiplication problems. It was ensured that the participants did not feel as being evaluated. Similar to the threat group, the arithmetic task was presented on a laptop with a brief rest session. However, unlike the threat group, participants resumed the task without any reminders. They were simply told that they were to continue with the task as in the first session and attempt their best.

Experimental Tasks. Two experimental tasks (i.e., a computer-administered test and a paper-and-pencil test) in the mathematical domain were presented. The computer-administered task was an arithmetic problem verification task, performed on a HP Laptop. The task was presented on a 1366 x 768 resolution screen in a 42-point Courier New Font. Written instructions (identical to those given by the experimenter) were displayed on the screen with example problems. Problems were displayed horizontally in the centre of the screen in a standard arithmetic format « a x b = c ». The symbols and numbers were separated by spaces equal to the width of one character. At the beginning of each trial, an asterisk (*) was displayed in the centre of the screen as a fixation point for 750-ms, followed by the arithmetic equation. Participants were instructed to press the « L » or « S » keys on an AZERTY keyboard to indicate whether the equation was true or false. The equation remained on the screen until the participant responded. The E-prime software controlled stimulus display, response recording, and collected response times with 1-ms accuracy. Before the experiment, participants practiced on 15 problems. The experimenter gave feedbacks on practice trials. Then, followed the experimental trials presented in two blocks of 144 problems each, separated by a brief break. The two blocks were identical; they included equal numbers of true and false problems and of problems from each even-operand condition.

The computer-administered test was followed by the paper-and-pencil test. This was an Arithmetic Fluency Test (French, Ekstrom & Price, 1963) where participants were asked to perform three sets of problems, each including basic additions, subtractions, and multiplications under a restricted time-frame (i.e. each of the three sets was time-bound for two minutes). All participants then completed the French version of the Mill-Hill Vocabulary Scale (MHVS; Deltour, 1993; Raven, 1951) which was administered intentionally after debriefing the participants about the goal of the study to eliminate any threat effects on these baseline measures of older adults' verbal abilities. The experimental session ended with the experimenter answering any further questions posed by the participants who were eventually thanked for their participation.

Results

Results are reported in three main parts. We examined the effects of stereotype threat on participants' performance first for false problems, second for true problems, and finally on the paper-and-pencil test. Two young participants in the stereotype threat group were excluded[1] from the analyses because they made 52% and 53% errors, respectively.

Effects of age-based stereotype threat on false problems performance.

Mean response latencies and percentages of errors on false problems were analyzed with 2 (Age: young, older adults) x 2 (Group: threat, control) x 2 (Parity: match, mismatch problems) mixed-design ANOVAs, with Age and Group as the between-participants factors (see means in Table 2 and summary of significant effects on latencies in Table 3)

[1] The exclusion/inclusion of these participants did not impact the results of any of the reported analyses.

Table 2. Young and Older Adults' Mean Response Latencies (in ms) and Percentages of Errors (in parentheses) on False and True Problems in the Threat and Control groups.

	Young Adults			Older Adults		
Problems	Threat	Control	*Means*	Threat	Control	*Means*
			False Problems			
Match	3940 (11.4)	3871 (10.3)	3905 (10.8)	4239 (6.0)	3891 (6.0)	4065 (6.0)
Mismatch	3511 (7.0)	3345 (5.7)	3428 (6.3)	4159 (3.0)	3461 (2.3)	3810 (2.7)
Means	3725 (9.2)	3608 (8.0)	3667 (8.6)	4199 (4.5)	3676 (4.2)	3938 (4.4)
PE	429** (4.9)	526** (4.6)	477 (4.5)	80 (3.3)	430** (3.7)	255 (3.3)
			True Problems			
Large	4548 (17.7)	4461 (15.9)	4504 (16.8)	5314 (11.3)	4550 (9.9)	4932 (10.6)
Small	4262 (11.7)	4218 (10.1)	4240 (10.9)	4673 (8.2)	4197 (8.7)	4435 (8.5)
Means	4405 (14.7)	4339 (13)	4372 (13.8)	4993 (9.7)	4373 (9.3)	4683 (9.6)
PSE	286** (6.0)	243** (5.8)	264 (5.9)	641** (3.1)	353** (1.2)	497 (2.1)

*Note: PE (Parity Effects) = Match - Mismatch; PSE (Problem Size Effects) = Large – Small; * $p < .05$; ** $p < .01$.*

Table 3. Statistics of Significant Effects in Experiment 1 for Latencies on False and True Problems.

Effects	*df*	*MSe*	*F*	η_p^2
False Problems				
Group	1,154	4,039,622	15.22***	0.09
Age	1,154	2,888,820	10.88***	0.07
Parity	1,154	10,583,599	165.39***	0.52
Age x Group	1,154	1,026,968	6.13*	0.04
Age x Parity	1,154	979,619	15.31***	0.09
Group x Parity	1,154	984,149	15.38***	0.09
Age x Group x Parity	1,154	316,977	4.95*	0.03
True Problems				
Group	1,154	4,625,682	12.48***	0.08
Age	1,154	3,830,163	10.34**	0.06
Problem Size	1,154	11,447,560	137.30***	0.47
Age x Group	1,154	3,029,754	8.18**	0.05
Age x Problem Size	1,154	1,068,068	12.81***	0.08
Group x Problem Size	1,154	538,521	6.46*	0.04
Age x Group x Problem Size	1,154	296,988	3.56	0.02

Note: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$

Young adults were 270 ms faster than older adults. Participants were 320 ms slower in the threat group than in the control group, and overall young and older adults were 366 ms faster on mismatch problems than on match problems. The significant Age x Group interaction showed that young adults tested in the threat group and in the control group were equally fast ($F<1$) while older adults were 523 ms slower in the threat group than in the control group ($F(1,154) = 20.60$, $MSe = 5,467,965$, $\eta^2_p = 0.12$). Most interestingly, the Age x Group x Parity interaction was significant, and further contrasts revealed that the effects of threat were larger on older adults' latencies while rejecting mismatch problems (threat – control = 698 ms; $F(1,154) = 31.22$, $MSe = 9,740,590$, $\eta^2_p = 0.17$) than when rejecting match problems (threat –

control = 348 ms; $F(1,154) = 8.56$, $MSe = 2,420,340$, $\eta^2_p = 0.05$). In contrast, young adults in the threat and control groups were equally fast for both match and mismatch problems ($Fs<1$). In other words, stereotype threat only affected older adults' verification times, and this on mismatch problems more than on match problems. Also, performance differences were found between young and older adults only in the threat group, on both match problems (299 ms; $F(1,154) = 6.15$, $MSe = 1,739,368$, $\eta^2_p = 0.04$) and mismatch problems (648 ms; $F(1,154) = 26.25$, $MSe = 8,188,926$, $\eta^2_p = 0.15$). In contrast, young and older adults in the control group had comparable performance on both match and mismatch problems ($Fs<1$). Parity effects (see Figure 1) were significant in young adults, both in the threat group (match – mismatch = 429 ms; $F(1,154) = 54.70$, $\eta^2_p = 0.26$) and in the control group (match – mismatch = 526 ms; $F(1,154) = 86.40$, $\eta^2_p = 0.36$). However, no parity effects were found in older participants tested in the threat group (match – mismatch = 80 ms; $F<2$) in contrast to significant parity effects in older participants tested under the control group (match – mismatch = 430 ms; $F(1, 154) = 57.71$, $\eta^2_p = 0.27$).

Analyses of errors revealed that young adults made more errors than older adults (8.6% *vs.* 4.3%), $F(1,154) = 25.10$, $MSe = 721$, $\eta^2_p = 0.14$. Participants made more errors on match than on mismatch problems (8.4% *vs.* 4.5%), $F(1,154) = 94.03$, $MSe = 1,217$, $\eta^2_p = 0.38$. No other effects came out significant on either latencies or errors.

Effects of age-based stereotype threat on true problems performance.

Participants' performance on true problems were analyzed with 2 (Age: young, older adults) x 2 (Group: threat, control) x 2 (Problem Size: small, large problems) mixed design ANOVAs, with Age and Group as the only between-participants factors (see means in Table 2 and summary of significant effects on latencies in Table 3).

Young adults were 311 ms faster than older adults. Participants in the control group were 342 ms faster than participants in the threat group, and were 380 ms slower on large problems than on small problems. The significant Age x Group interaction showed that young participants in the threat and control groups were equally fast (F<1) whereas older adults in the threat group were 619 ms slower than older adults in the control group, $F(1,154) = 20.70$, $MSe = 7,670,959$, $\eta^2_p = 0.12$. Most interestingly, the Age x Group x Problem Size interaction was significant, and contrasts revealed that the effects of threat were larger on older adults' latencies while verifying true large problems (threat – control = 764 ms; $F(1,154) = 24.65$, $MSe = 11,650,248$, $\eta^2_p = 0.14$) than while verifying true small problems (threat – control = 476 ms; $F(1,154) = 12.85$, $MSe = 4,520,103$, $\eta^2_p = 0.08$). In contrast, young adults in the threat and control groups were equally fast for both large or small problems (Fs<1). In other words, stereotype threat only affected older adults' verification times, and this on large problems more than on small problems. Also, performance differences were found between young and older adults only in the threat group, on both large problems (766 ms; $F(1,154) = 24.20$, $MSe = 11,437,912$, $\eta^2_p = 0.14$) and small problems (411 ms; $F(1,154) = 9.35$, $MSe = 3,289,539$, $\eta^2_p = 0.06$). In contrast, young and older adults in the control group had comparable performance on both large and small problems (Fs<1). Problem-size effects (see Figure 1) were of the same magnitudes in young adults tested in the threat group (large – small = 286 ms; $F(1,154) = 18.60$, $\eta^2_p = 0.11$) and in the control group (large – small = 243 ms; $F(1,154) = 14.19$, $\eta^2_p = 0.08$). However, larger problem-size effects were found for older adults in the threat group (large – small = 641 ms; $F(1,154) = 98.55$, $\eta^2_p = 0.39$) compared to the control group (large – small = 353 ms; $F(1,154) = 29.92$, $\eta^2_p = 0.16$).

Analyses of errors revealed that young adults made more errors than older adults (13.8% *vs.* 9.5%), $F(1,154) = 23.99$, $MSe = 745$, $\eta^2_p = 0.14$. Participants made more errors on large than on small problems (13.8% *vs.* 9.7%), $F(1,154) = 75.30$, $MSe = 1,286$, $\eta^2_p = 0.33$. No other effects came out significant on either latencies or errors.

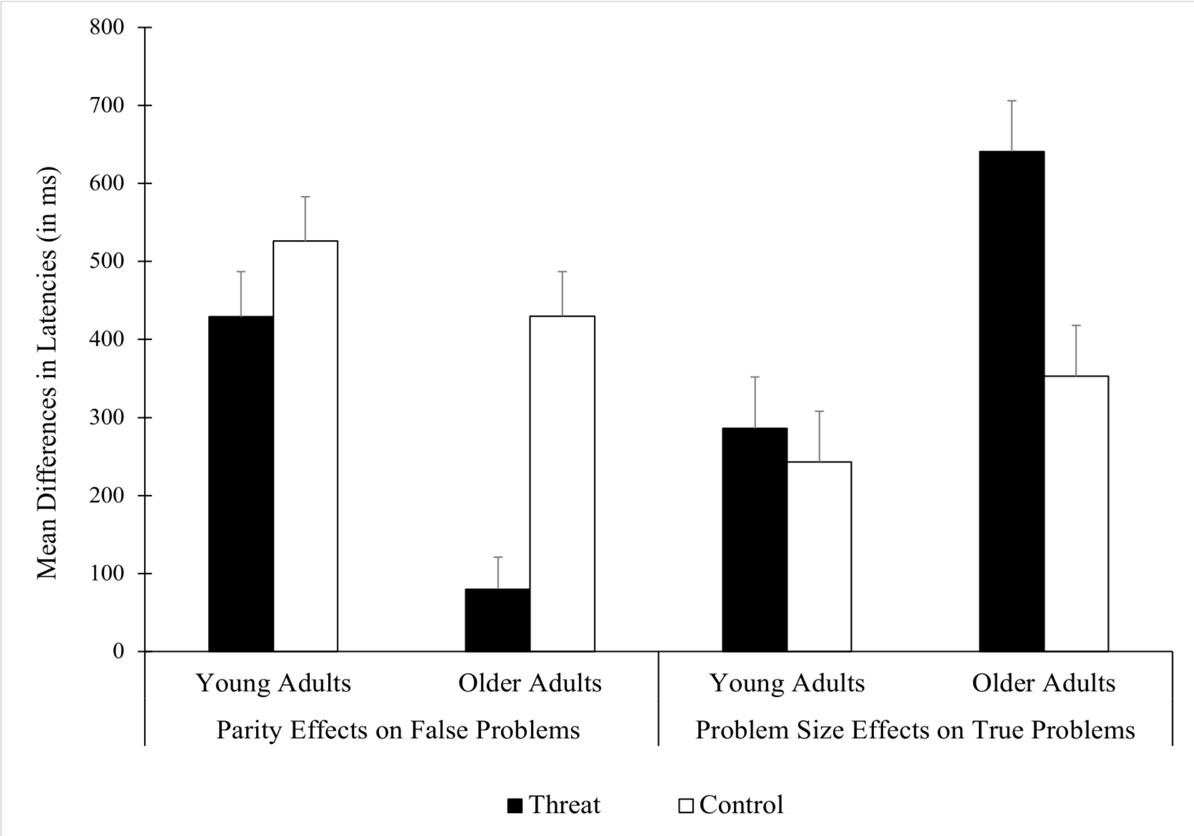


Figure 1. Parity Effects (mean differences in verification latencies between False Match and Mismatch problems) and Problem Size Effects (mean differences in verification latencies between True Small and Large problems) in Young and Older Adults under Threat and Control Groups. Error Bars represent 95% Confidence Intervals.

Effects of age-based stereotype threat on participants' performance in the paper-and-pencil arithmetic test.

Older adults' performance on the arithmetic fluency test (French et al., 1963) was analyzed with a 2 (Age: young, older adults) x 2 (Group: threat, control) between-participants ANOVA. Results showed a main effect of Age on arithmetic fluency scores, $F(1,154) = 367.4$, $MSe = 105{,}073$, $\eta^2_p = 0.71$ (i.e., older adults scored higher than young adults on arithmetic fluency test, 79.6 *vs.* 28.1). Interestingly, the Age x Group interaction was significant, $F(1,154) = 12.62$, $MSe = 3{,}609$, $\eta^2_p = 0.08$. This was the result of better performance in older adults tested in the control group than older adults tested in the threat group (86.9 *vs.* 72.5; $F(1,154) = 14.55$, $MSe = 4{,}161$, $\eta^2_p = 0.09$) and of equal performance in both groups of young adults ($Fs<2$).

Experiment 2

In Experiment 1, we found decreased performance in older adults only, under stereotype threat condition relative to control condition. More specifically, we found increased problem size effects and decreased parity-violation effects. Increased problem size effects were the result of older adults executing the calculation strategy more slowly on large problems under stereotype threat. In other words, as we hypothesized, stereotype threat led older adults to execute strategies more poorly under stereotype threat condition, especially on large problems. Decreased parity-violation effects under stereotype threat resulted from older adults taking longer to verify false mismatch problems. This likely happened because participants did not use as often and as systematically the fast plausibility checking strategy (based on the discrepancy between the parity of the proposed and correct answers) on false mismatch problems but used the slower exact calculation strategy under stereotype threat.

One limitation of Experiment 1 regarding the strategy hypothesis is that strategy use and strategy execution were not investigated directly. They were inferred from patterns of participants' performance as a function of stimulus features (problem size, parity). Such an inference runs the risk of possible other sources of differences in magnitudes of problem size and parity violation effects between control and threat conditions. Also, strategy use and strategy execution cannot be assessed independently, as differences in participants' performance between control and threat conditions result from differences in both strategy use and strategy execution.

Experiment 2 aimed at testing the strategy hypothesis more directly by using a task where strategy use is not inferred from participants' performance but observed with external behavioral evidence (i.e., calculating out loud) and where it is possible to assess ABST effects on strategy use and strategy execution independently. We asked participants to accomplish computational estimation tasks (i.e., finding the best estimates to two-digit multiplication problems, like 32x79). Numerous previous studies used this task to investigate strategic aspects of arithmetic problem solving performance and age-related differences therein (e.g., LeFevre, Greenham, & Waheed, 1993; Hinault & Lemaire, 2017; Lemaire & Leclere, 2014a). We also used the Choice/No-Choice method proposed by Siegler and Lemaire (1997) as used in many studies since then. This method enables to investigate strategy use and strategy selection (in the Choice condition) independently of strategy execution (in the No-choice condition). Asking participants to execute the cued strategy in the No-choice condition controls for strategy selection artifacts, thereby yielding unbiased estimate of strategy execution. Because we wanted to examine strategy execution independently of strategy selection, ABST effects on performance were assessed in the choice condition and by no-Choice condition, separately. This sequence was also important to prevent strategy choices in the choice condition from being affected by carry-over effects from strategy execution in the no-choice condition.

In the Choice condition, participants were asked to choose between two strategies (i.e., rounding-down or rounding-up strategies) to provide estimates of two-digit multiplication problems (e.g., 63x82). Participants were instructed to use the better of the two strategies on each problem to provide the best estimate in the Choice condition, and to execute the cued (rounding-down or rounding-up) strategy on each problem in the No-Choice condition. The hypothesis that ABST influences strategy use predicted that participants use the better strategy less often in the threat than in the control group. Independent measures of strategy used enabled us to conduct a mediation analysis with strategy use as the underlying mechanism of ABST effects on arithmetic performance. We predicted a substantial attenuation of the relation between ABST and arithmetic performance, after statistically controlling for better strategy use. Also, and most originally, we tested the prediction of increased strategy repetitions (i.e., the tendency to repeat a strategy over consecutive problems even when the better strategy is different on two consecutive problems) in the threat relative to the control group, thereby providing converging evidence on less efficient use of the better strategy.

Analyses of performance in the No-Choice condition enabled us to examine ABST effects on strategy execution uncontaminated by effects on strategy use. In the No-Choice condition, participants did not select strategies on each problem but were asked to execute either the poorer or the better strategy. Previous studies (e.g., Hinault & Lemaire, 2017; Lemaire & Hinault, 2014; Hinault, Dufau, & Lemaire, 2014) found that participants' performance is slower when asked to execute the poorer strategy than when asked to execute the better strategy. Comparing poorer strategy effects (i.e., differences in performance between better and poorer strategy trials) in the threat and control groups enabled us to test an additional prediction derived from the hypothesis that ABST will impair older adults' execution. We expected that larger poorer-strategy effects (i.e., differences in participants' performance while executing the poorer and the better strategies) in the threat group than in the control group.

Method

Participants.

Participants were 56 older adults who were all volunteers, recruited from the community. Prior to the experiment, all older adults accomplished the Mini Mental-State Examination (MMSE, Folstein et al., 1975), and none obtained a MMSE score < 27. Participants also completed a French version of the Mill-Hill Vocabulary Scale (MHVS; Deltour, 1993; Raven, 1951) to assess their verbal abilities. Participants' characteristics are summarized in Table 4. A-priori power analysis (G*Power; Faul et al., 2007) was conducted to determine the sample size needed to detect our key between-within interactions: The 2(Group: threat, control) x 2(Trials: repeated, unrepeated) interaction to assess the influence of ABST on strategy repetitions in the Choice condition, and the 2(Group: threat, control) x 2(Trial: poorer-strategy, better-strategy) interaction to assess the influence of ABST on older adults' execution in the No Choice condition. With an assumed *d* of .52 (Lamont et al., 2015), our mixed design could achieve 80% power with as few as 46 participants, given a modest correlation ($r = .30$) between our repeated measures. We stopped our participants recruitment to attain 28 participants per group, resulting in a sample of 56 older adults.

Table 4. Participants' Characteristics in the Threat and Control Groups (Expt. 2).

	Threat *M(SD)*	Control *M(SD)*	*p*
N (Females)	28 (22)	28 (19)	--
Age (Years)	73.2 (5.4)	72.6 (5.5)	0.69
Education (Years)	13.1 (2.5)	12 (1.6)	0.07
MMSE	29 (0.8)	28.9 (0.8)	0.09
MHVS	28 (3.1)	26 (5.4)	0.07

Stimuli.

The stimuli for the computational estimation task were 64 two-digit multiplication problems presented in a standard form (e.g., 32 x 67). Based on the size of the unit digits, half the problems were so-called homogenous problems, and half were so-called heterogeneous problems. Unit digits of both operands were either smaller than five (e.g., 32 x 63) or larger than five (e.g., 38 x 69) for homogeneous problems. Heterogeneous problems had unit digits larger than five in the first operand and smaller than five in the other operand (e.g., 49 x 62) for half the problems and the reverse (e.g., 42 x 58) for the other problems. Half the homogeneous and heterogeneous problems were best estimated (i.e., they yielded closest products from the correct products) with the rounding-down strategy (e.g., a homogeneous problem, like 83 x 21, and a heterogeneous problem, like 86 x 21, are best estimated with the rounding-down strategy, doing 80 x 20 = 1600). The other problems were best estimated with the rounding-up strategy (e.g., a homogeneous problem, like 78 x 39, and a heterogeneous problem, like 72 x 39, are best estimated with the rounding-up strategy, doing 80 x 40 = 3200).

We tested homogeneous and heterogeneous problems because previous studies found that participants select the best rounding strategy more easily on homogenous than on heterogeneous problems (e.g., LeFevre, Greenham, & Waheed, 1993; Lemaire, Arnaud, & Lecacheur, 2004). Homogeneous and heterogeneous problems had comparable exact products when solved with each rounding strategy (*Means* were 2426 and 2408 for homogeneous and heterogeneous problems, respectively).

Finally, following previous studies on strategy repetition (e.g., Lemaire & Brun, 2016, Lemaire & Leclère, 2014a, b), we tested whether participants have a tendency to repeat the same strategy on consecutive trials more often in the threat than in the control condition. Two types of trials were tested, one for which it was appropriate to repeat the same strategy across two consecutive trials in order to select the best strategy and one where it was not appropriate to repeat the same strategy. Thus, half the trials were repeated-strategy trials (i.e., two successive problems were best estimated with the same strategy), and half were unrepeated-strategy trials (two successive problems were best estimated using different strategies).

Given well-known effects in the domain of mental arithmetic (see Kadosh & Dowker, 2015; Gilmore et al., 2018, for overviews), the following factors were controlled: (a) no operand had zero or five as unit digit, (b) digits were not repeated within operands (e.g., 33 x 42) and neither were the operands for any given problem (e.g., 42 x 42), (c) digits were not repeated in the same unit or decade positions across operands (e.g., 62 x 67), (d) no reverse order of operands was used (e.g., 56 x 23 vs. 23 x 56), (e) the first operand was larger than the second operand in half the problems, and smaller in the other problems, and (f) the operand with the smallest unit digits was in the left position in half the problems (e.g., 42 x 36) and in the right position in the other problems (e.g., 23 x 41).

Procedure.

Participants were tested in one session that lasted approximately 60 minutes. Each participant was randomly assigned to either a Threat group or a Control group. The stereotype threat manipulation was the same as in Experiment 1 and was implemented exactly the same way (same instructions).

The computational estimation task was performed on a HP Laptop. The task was presented on a 1366 x 768 resolution screen in a 42-point Courier New Font. Written instructions (identical to those orally given by the experimenter) were displayed on the computer screen with example problems. Problems were displayed horizontally in the centre of the screen in a standard arithmetic format « *a* x *b* ». The symbols and numbers were separated by spaces equal to the width of one character. Participants solved all problems twice, first in a Choice condition (to investigate strategy choices), second in a No-Choice condition (to study strategy execution).

In the Choice condition, participants were asked to try to use the best strategy they could to obtain the best estimate for each problem. The best strategy was defined as the strategy that yields the best estimate (i.e., closest from correct product). Participants were told that they could choose between the rounding-down strategy (i.e. rounding the two operands down to the nearest decades; 80 x 20 to estimate 86 x 21) or the rounding-up strategy (i.e. rounding the two operands up to the nearest decades; 90 x 30 to estimate 87 x 26), and that no other strategy (e.g., mixed rounding strategy; rounding the first operand down and the second operand up to the closest decades; 80 x 30 to estimate 84 x 29) could be used. By excluding the use of mixed-rounding strategy, we made strategy choice harder for the participants (especially on heterogeneous problems, where it is easier to use a mixed-rounding strategy over rounding-down or rounding-up strategy). Since previous research on computational estimation strategies observed that participants use the mixed-rounding strategy (e.g., LeFevre et al., 1993; Lemaire, Arnaud, & Lecacheur, 2004), it may be that threat-related differences on strategy use do not emerge if participants could use mixed-rounding strategy. This may be a result of all participants (in both threat and control groups) making use of the mixed-rounding strategy to solve all the problems. Prohibiting the use of this strategy offered us an opportunity to investigate strategy selection process under threat by varying the problem features (e.g., size of the unit digits) and then assessing participants' difficulty to find the best strategy on each problem. Participants were instructed to give their estimates out loud and to indicate which strategy (rounding-down or rounding-up) they used on each problem. The estimate and the strategy were recorded by the experimenter on each trial.

In the No-Choice condition, to make sure participants executed the required strategy on each problem, they had to execute strategies and provide estimates out loud for the cued strategy. The better of the two strategies was cued on half the problems (e.g., rounding up was cued for 56 x 78), and the poorer strategy was required on the other problems (e.g., rounding down was cued for 57 x 68). We cued the better or poorer strategy on half the problems, as previous studies reported so-called poorer-strategy effects. In these effects, participants are slower when they execute a poorer strategy to solve a problem than when they execute a better computational estimation strategy (e.g., Hinault & Lemaire, 2017; Hinault, Badier, Baillet, & Lemaire, 2017; Hinault, Lemaire, & Phililips, 2016; Hinault, Lemaire, & Touron, 2016; Hinault et al., 2014; Lemaire & Hinault, 2014; Lemaire et al., 2004). The experimenter recorded participants' estimate on each problem.

For both the Choice and No-Choice conditions, each problem remained on the screen until participants responded. Following procedure used in previous studies (e.g., Hinault & Lemaire, 2017, Lemaire et al., 2004) on computational estimation, a timer was started at the beginning of the problem presentation and ended when the experimenter clicked the mouse button, which happened as quickly as possible when participants started to state their response orally. Like in Experiment 1, the E-prime software controlled stimulus display, response recording, and collected latency data with 1-ms accuracy. Before each condition, two example problems were explained, and the experimenter gave feedback when necessary. Following Siegler and Lemaire (1997), participants were first tested under the Choice condition and then under the No-Choice condition, with a brief period of 5-10 minutes break between the two conditions, so that strategy use in Choice condition was not contaminated by recent execution of strategies in the No-Choice conditions.

Results

Results are reported in two main parts. We examined the effects of age-based stereotype threat first on strategy selection and participants' performance in the Choice condition, and second on strategy execution in the No-Choice condition.

Effects of Age-Based Stereotype Threat on Strategy Selection and Performance in the Choice condition.

Strategy Selection. We first conducted a 2 (Group: threat, control) x 2 (Problem: homogeneous, heterogeneous) mixed design ANOVA on mean percentages of use of the better strategy, with Group as the between-participants factor (see means in Table 5).

Table 5. Mean Percentages of Use of Better Strategy, Mean Solution Latencies, and Absolute Percent Deviations (in parentheses) on Homogeneous and Heterogeneous Problems in the Choice Condition for Threat and Control Groups.

Problems	Threat	Control	*Means*
	Percentages of Use of Better Strategy (%)		
Homogeneous	67	94	80
Heterogeneous	55	70	62
Means	61	82	71
	Mean Solution Latencies in ms (and percent deviations)		
Homogeneous	4088 (11.3)	5526 (1.9)	4807 (6.6)

Heterogeneous	5091 (9.4)	5730 (7.7)	5410 (8.6)
Means	4590 (10.3)	5628 (4.8)	5109 (7.6)

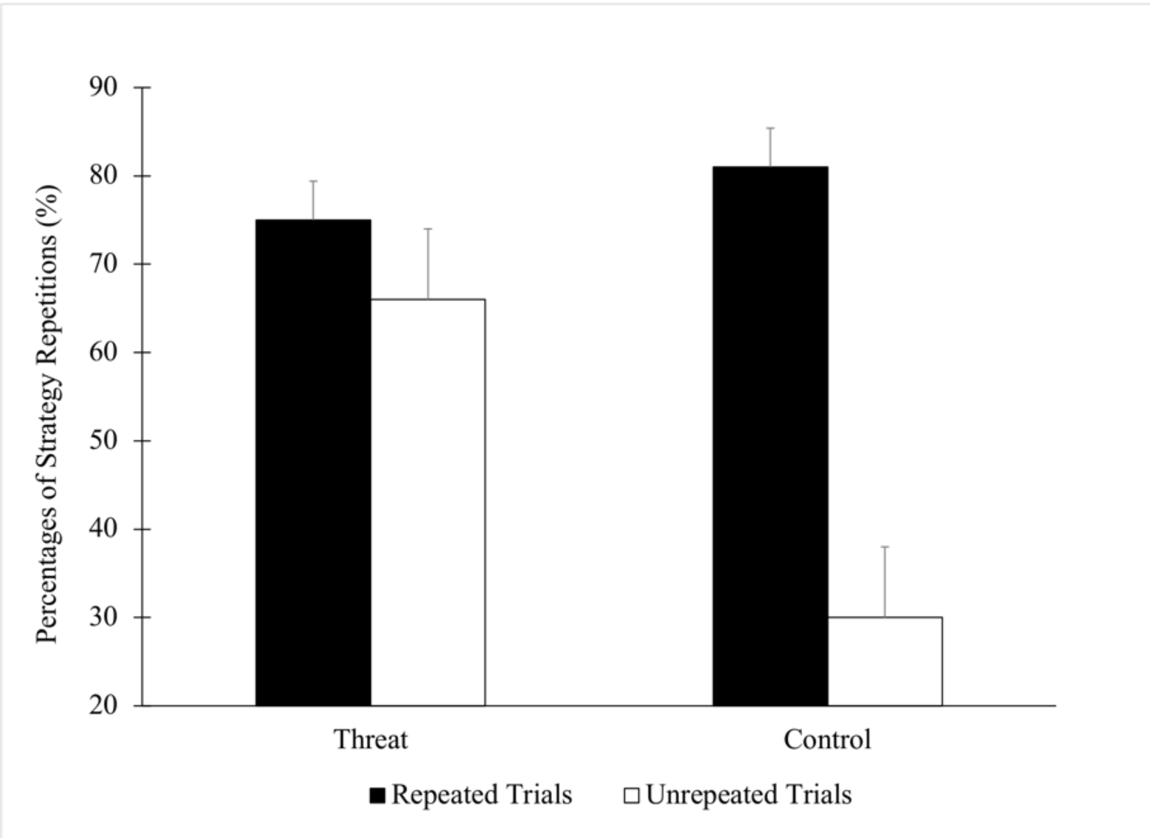
Results showed that overall older adults selected the better strategy more often on homogeneous than on heterogeneous problems, $F(1,54) = 113.4$, $MSe = 8{,}984$, $\eta^2_p = 0.68$) and in the control than in the threat group, $F(1,54) = 49.2$, $MSe = 6{,}246$, $\eta^2_p = 0.48$). The significant Group x Problem interaction, $F(1,54) = 10.68$, $MSe = 846$, $\eta^2_p = 0.16$, showed that the effect of threat was larger on homogeneous problems (control – threat = 27%; $F(1,54) = 51.8$, $MSe = 9{,}922$, $\eta^2_p = 0.49$) than on heterogeneous problems (control – threat = 15%; $F(1,54) = 24.1$, $MSe = 3{,}418$, $\eta^2_p = 0.31$). Thus, older adults selected the better strategy less often under stereotype threat than under control group, and this effect was larger on homogeneous than on heterogeneous problems.[2]

Sequential effects during strategy selection. We conducted a 2 Group (threat, control) x 2 Trial (repeated, unrepeated) design ANOVA on mean percentages of strategy repetition, with Group as the only between-participants factor. Strategy repetition was coded 1 if a participant used the same strategy on two consecutive problems and 0 otherwise.

Participants repeated the same strategy across two successive problems more often on repeated trials (78%) than on unrepeated trials (48%, $F(1,54) = 110.8$, $MSe = 24{,}213$, $\eta^2_p = 0.67$). Moreover, participants in the threat group repeated strategies across trials more often (71%) than participants in the control group (55%, $F(1,54) = 22.9$, $MSe = 3{,}105$, $\eta^2_p = 0.30$).

[2] Examination of individual data for strategies used by individuals revealed that there were two groups of participants in the stereotype threat group, single-strategy users (who used only one strategy on more than 80% of the problems) and dual-strategy users (who used both strategies). To test the possibility that stereotype threat affected strategy selection because of the presence of single-strategy users in the threat group, we ran a 2 (Group: Threat, Control) x 2 (Problems: Homogeneous, Heterogeneous) design ANOVA exclusively for dual-strategy users on mean percentages of use of the better strategy. Results confirmed that threat still affected older adults' strategy use. Older adults in the threat group selected the better strategy less often than older adults in the control group (68% *vs*. 82%; $F(1,40) = 14.5$, $MSe = 1{,}871$, $\eta^2_p = 0.27$).

Most importantly, the Group x Trial interaction (Figure 2), $F(1,54) = 55.4$, $MSe = 12{,}118$, $\eta^2_p = 0.51$, showed larger rates of strategy repetitions across trials in the control group than in the threat group for repeated trials, (81% vs. 75%; $F(1,54) = 4.5$, $MSe = 489$, $\eta^2_p = 0.08$), and the reverse for unrepeated trials, (30% vs. 66%; $F(1,54) = 46.8$, $MSe = 17{,}839$, $\eta^2_p = 0.46$).



Fi [illegible] o successive problems were of the same type) and Unrepeated Trials (when two successive problems were of different types) in the Threat and Contol Groups. Error Bars represent 95% Confidence Intervals.

Participants' performance. We carried out 2 (Group: threat, control) x 2 (Problem: homogeneous, heterogeneous) mixed design ANOVAs on mean solution latencies and absolute percent deviations, with Group as the only between-participants factor (See means in Table 5).

Participants were faster on homogeneous than on heterogeneous problems, $F(1,54) = 39.3$, $MSe = 10,211,828$, $\eta^2_p = 0.42$. Also, participants in threat group were faster than participants in the control group, $F(1,54) = 7.7$, $MSe = 15,090,963$, $\eta^2_p = 0.13$. The Group x Problem interaction revealed significant differences between the threat and the control groups on homogeneous problems only, $F(1,54) = 14.09$, $MSe = 28,952,692$, $\eta^2_p = 0.21$. In other words, participants in the threat group were 1438 ms faster than participants in the control group on homogeneous problems but equally fast on heterogeneous problems ($F<3.0$). Interestingly, analyses on absolute percent deviations revealed that participants' estimates deviated more from the correct products in the threat group than in the control group (mean difference = 5.6; $F(1,54) = 16.46$, $MSe = 434$, $\eta^2_p = 0.23$) and more on heterogeneous than on homogeneous problems (mean difference = 1.9; $F(1,54) = 6.30$, $MSe = 106$, $\eta^2_p = 0.10$). The Group x Problem interaction revealed significant differences between the threat and the control groups on homogeneous problems only, $F(1,54) = 43.65$, $MSe = 1,253$, $\eta^2_p = 0.45$). In other words, participants in the threat group deviated 9.5% more than participants in the control group on homogeneous problems. Percent deviations for participants in both the threat and the control groups were comparable on heterogeneous problems ($F<1$).

Mediation analyses. To determine whether percentages of use of the better strategy mediated effects of stereotype threat on absolute percent deviations, simple mediation analyses were conducted on absolute percent deviations in homogeneous and heterogeneous problems pooled together. Using the PROCESS macro for IBM SPSS (10,000 bootstrapped resamples; Model 4; Hayes, 2013), we regressed the deviations on Group (dummy coded: 0 = Control and 1 = Threat) and entered mean percentages of use of better strategy as the mediator. Participants in the threat group used the better strategy less often than those in the control group (a = -21.2). As can be seen in Figure 3, the more often participants selected the better strategy the smaller the deviation (b = -0.29). The confidence interval of the indirect effect through percentages of use of the better strategy did not include zero (ab = 6.08 (1.7), 95%CI = [3.5, 10.2]). In other words, there was no evidence that stereotype threat influenced percent deviations independent of its effect on percentages of use of the better strategy (c' = -0.9, p = 0.6). Percentages of use of the better strategy was thus a significant mediator of stereotype threat on percent deviations.

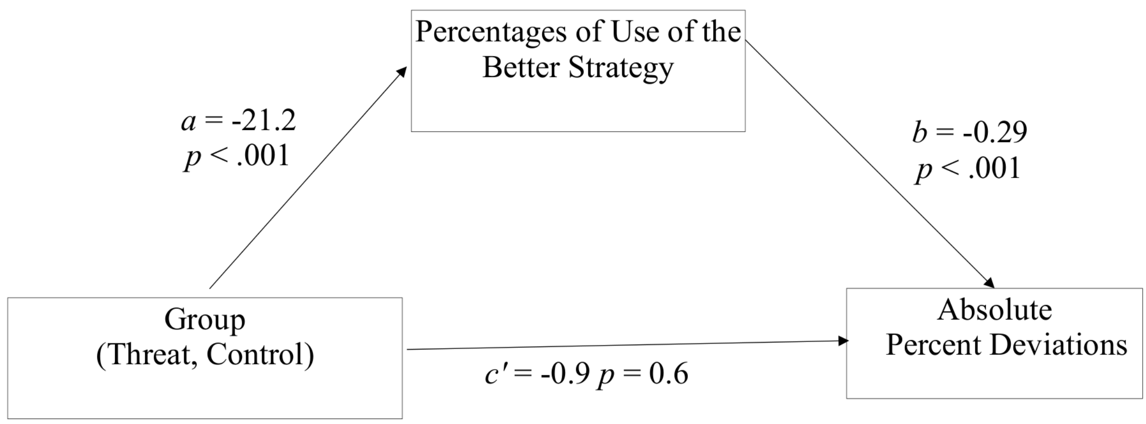


Figure 3. Simple Mediation Model for the Percentages of Use of the Better Strategy on Absolute Percent Deviations.

Effects of Age-Based Stereotype Threat on Strategy Performance in the No-Choice Condition.

We conducted 2 (Group: threat, control) x 2 (Problem: homogeneous, heterogeneous problems) x 2 (Trial: poorer-strategy, better-strategy) mixed-design ANOVAs on mean solution latencies and absolute percent deviations, with Group as the only between-participants factor (see means in Table 6).

Table 6. Older Adults' Mean Solution Latencies (in ms) and Absolute Percent Deviations (in parentheses) on Homogeneous and on Heterogeneous Problems when executing a Poorer or a Better Strategy in the Threat and Control groups.

	Threat			Control		
Problems	Poorer Strategy	Better Strategy	*Differences*	Poorer Strategy	Better Strategy	*Differences*
Homogeneous	6071 (1.1)	4758 (2.7)	1313 (-1,6)	5120 (1.1)	4748 (1.2)	372 (-0.1)
Heterogeneous	6136 (1.1)	4864 (1.8)	1272 (-0.7)	5032 (1.2)	4794 (0.9)	238 (0.3)
Means	6104 (1.1)	4811 (2.2)	1293 (-1.1)	5076 (1.2)	4771 (1.0)	305 (0.2)

Note: Differences = Poorer-Better.

Participants were slower while executing the poorer-strategy relative to while executing the better strategy on a given problem (4791 ms *vs*. 5590 ms; $F(1,54) = 161.07$, $MSe = 35,745,662$, $\eta^2_p = 0.75$). The significant Group x Trial interaction revealed significant differences between participants' latencies in the threat and control groups on poorer-strategy trials only, $F(1,54) = 8.8$, $MSe = 14,781,615$, $\eta^2_p = 0.14$. In other words, older adults were slower on poorer-strategy trials when tested under stereotype threat than when tested under control group (*Mean difference* = 1,028 ms). Older adults had comparable performance under threat and control groups on better-strategy trials (*Mean difference* = 40 ms; $F<1.0$). The Group x Trial x Problem interaction was not significant, $F<1.0$.

Analyses of mean absolute percent deviations revealed a significant Group x Trial interaction. This resulted from differences between the threat and the control groups being significant on better-strategy trials only, $F(1,54) = 4.8$, $MSe = 22$, $\eta^2_p = 0.08$. Participants' estimates deviated 1.3% more in the threat group than in the control group when asked to execute the better strategy on a given problem.

Experiment 3

ABST effects on older adults' cognitive performance have always been documented using a between-participants design. This makes it impossible to determine whether stereotype threat can be manipulated within the same individual, and to examine how and to what extent an individual's cognitive performance changes from a baseline, reduced threat to a threat condition. Experiment 3 therefore aimed at testing whether a pretest-posttest design is an appropriate method to examine threat sensitivity in older adults such that, as assumed by Steele (2010), threat can be elicited within the same individual just by changing the situational aspect of the presented task, and so as to be able to determine whether some older individuals are more sensitive than others to threat. We tested each individual before and after ABST instructions in the experimental group, and before and after control instructions in the control group. Differences in performance between pre-test and post-test under ABST condition, relative to corresponding differences under control, no-threat condition yielded the type of data that enabled us to examine individuals' sensitivity to threat, above and beyond mere test-retest effects. Further, to ensure that our threat and control groups were comparable, we collected measures of domain-general processing resources (i.e., processing speed) and of domain-specific arithmetic resources (i.e., arithmetic fluency) for all our participants.

Brubaker and Naveh-Benjamin (2018) recently used a within-participants design and demonstrated ABST effects on both item and associative memory in the same individuals tested under nonthreat and threat conditions. Here, we used a pretest-posttest manipulation in another cognitive domain (arithmetic) and controlled for test-retest effects to examine changes in strategic variations within the same individuals. Also, Experiment 3 offered the opportunity to examine relations among variations in magnitudes of ABST effects and individual characteristics such as domain-general processing (speed) and domain-specific (arithmetic fluency) resources.

Method

Participants.

Forty-eight older adults participated. They were all volunteers recruited from the community. Prior to the experiment, all participants accomplished the MMSE (Folstein et al., 1975) and the French version of the Mill-Hill Vocabulary Scale (MHVS; Deltour, 1993; Raven, 1951). None obtained an MMSE score < 27. Therefore, none were excluded. Likewise, participants' verbal fluency was quite high and similar in the Threat and Control groups. Participants' characteristics are summarized in Table 7. A-priori power analysis (G*Power; Faul et al., 2007) indicated that a sample size of 46 would be sufficient for our 2x2 mixed design to achieve at least 80% power to detect the critical 2(Group: threat, control) x 2(Session: pre-test, post-test) interaction. To calculate this sample size we used the same parameters as used in Experiment 2 (i.e, $d = 0.52$, and $r = 0.30$) for the same reasons.

Table 7. Participants Characteristics in Threat and Control Groups.

	Threat *(SD)*	Control (*SD*)	*p*
N (Females)	24 (19)	24 (13)	--
Age (Years)	72 (4.6)	75 (8.3)	0.12
Education (Years)	13 (1.9)	12 (1.6)	0.38
MMSE	29 (0.8)	29 (0.9)	0.27
MHVS	22.3 (5.6)	20 (5.6)	0.12

Processing speed measures. The WAIS-IV Digit symbol coding task (Codes) and the Choice reaction time task (Color Detection) were used to measure speed of processing. The WAIS-IV Digit symbol coding task (Wechsler, 1997) is a pencil-and-paper task that involves the presentation of a code table containing nine digit-symbol pairs at the top, and a series of digits below. Participants had two minutes to complete as many as possible blank spaces under each digit with the symbol corresponding to the digit. For each individual, we assessed the number of items correctly completed within two minutes.

The choice reaction time task consists of 30 trials for which a red or a black star was displayed in the middle of the laptop screen. Participants had to determine as fast as possible if the presented star was black or red by pressing a key on an AZERTY keyboard. Before beginning the task, participants completed a short training of 10 trials. On each trial latencies were collected as a measure of processing speed.

Arithmetic fluency. Participants' arithmetic fluency was assessed using the French Kit (French et al., 1963). They were asked to perform three sets of arithmetic problems, each including basic addition, subtraction, and multiplication problems, under a restricted time-frame (i.e. each of the three sets was time-bound for two minutes).

Stimuli.

The computational estimation task included 24 trials. Each of the 24 trials consisted of three consecutive two-digit multiplication problems. Each trial was followed by a series of four letters to avoid carry-over effects from one trial to the next (e.g., participants may repeat the same strategy on the last problem of a given trial and on the first problem of the next trial independently of which strategy is the best on each of these problems). Each series of three multiplication problems was composed of two priming problems followed by a target problem (e.g., priming problems, 78 x 27 and 48 x 79, followed by a target problem, 46 x 83). Priming problems were so-called homogeneous problems and target problems were so-called heterogeneous problems. Homogeneous (priming) problems were problems with the unit digits of both operands smaller or larger than 5 (e.g., 32 x 64 and 37 x 69). These problems were chosen as priming problems because previous works (e.g., LeFevre et al., 1993; Lemaire &

Brun 2014) showed that both young and older adults almost systematically select the better strategy on these problems. Heterogeneous (target) problems were problems with the unit digit of the operand smaller than 5 and the unit digit of the other operand larger than 5 (e.g., 43 x 69). They were selected as target problems as it is harder to select the better strategy on these problems for both young and older adults. This priming-target structure of trials followed several previous studies (Hinault et al., 2017; Lemaire & Leclère, 2014a,b) that revealed the appropriateness of this stimuli-structure to investigate age-related differences in strategy use and strategy execution. Here, we used this structure to test whether ABST changes (a) percentages of better strategy use and strategy performance on priming and target problems differently, and (b) how often participants tend to repeat the same strategy across priming and target problems. Strategy repetitions were tested on two types of trials, one for which it was appropriate to repeat the same strategy across two consecutive trials in order to select the best strategy and one where it was not appropriate to repeat the same strategy. Thus, half the trials were repeated-strategy trials (i.e., priming and target problems were best estimated with the same strategy), and half were unrepeated-strategy trials (priming and target problems were best estimated with different strategies). Following previous studies showing that strategy repetitions result from executive control mechanisms (e.g., Lemaire & Brun, 2014a,b; 2016, Lemaire & Leclère, 2014a,b), we studied the strategy repetition phenomenon here to determine whether executive control mechanisms can moderate ABST effect on strategy repetitions.

Finally, following previous findings in arithmetic (see Kadosh & Dowker, 2015; Gilmore et al., 2018, for overviews), the following factors were controlled: (a) no operands had zero or five as unit digits (e.g., 20 x 6, 25 x 63); (b) no digits were repeated within operands (e.g., 22 x 63); (c) the two operands of a given multiplication problem were never the same (e.g., 63 x 63); (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) rounded operands were never the same across two successive problems in a given trial (e.g., if one problem in a trial was 32 x 64, the next problem could not be 31 x 62).

Procedure.

The participants were individually tested in two sessions, each of which lasted approximately 45-60 minutes. The two sessions were at least 4—7 days apart. Session 1 was introduced as a set of tasks to be completed by all participants who intended to volunteer for the main experiment, scheduled for the next session. All participants in Session 1 completed the following tasks in the same order: Mill-Hill Vocabulary Scale, Codes, Color-Detection, and the French Kit. In Session 2, all participants completed the computational estimation task twice, as pre-test and post-test. In both pre- and post-tests, the two-digit multiplication problems were presented horizontally in a 84 point Bold Courier font (black color) on a 1366 x 768 resolution screen. The software E-Prime controlled the stimulus display and the latency collection. Participants were randomly assigned to either a control or a threat group, based on how the computational estimation post-test was presented.

*Control Group***.** The session was introduced as a study on problem solving in arithmetic, and participants were told that they were to perform a computational estimation task (e.g., 14 x 26) 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 (e.g., 70 x 40 to solve 78 x 43) and the rounding-up strategy as rounding both operands up to the nearest decades (e.g., 80 x 50 to solve 78 x 43). Further, the computational estimation task was presented as comprising of two blocks of 24 trials each. The two blocks were actually the pre-test and the post-test administered one after the other with a brief rest period of 3-5 minutes between the two tests. The pre-test and the post-test were the same computational estimation task with 24 trials each, with the following differences: (a) the order of operands of each problem was reversed (e.g., 78 x 27 in the pre-test was 27 x 78 in the post-test), (b) of the 24 trials, the last 12 trials of the pre-test were presented first in the post-test, and vice-versa, (c) the order of the trials were different in the two sessions but the sequence of the problems within each trial was the same.

Threat Group. The pre-test was presented as an optional task. That is, the experimenter pretended an urgent need to validate some arithmetic stimuli for an upcoming experiment and requested the participants to volunteer for this task in addition to the experiment for which they were recruited. All participants agreed to help the experimenter and were then presented the computational estimation task of 24 trials. The same instructions as in the control group were given (i.e., to provide an approximate answer to an arithmetic problem, like 14 x 26, using either a rounding-up or a rounding-down strategy). This followed a brief rest period of 3—5 minutes where the experimenter thanked the participants for their help and eventually invited them to participate in the experimental task, for which they were recruited. The post-test was then presented as a study of memory and attention abilities. Participants were told that they were going to perform a computational estimation task as earlier. However, there was a difference. The difference was explained as the computational estimation task being a memory test, comprising a set of arithmetic stimuli validated by experts in the memory domain to evaluate participants' cognitive abilities. Participants were specifically told that they were evaluated for their speed and accuracy. In addition, participants were cued that they could only successfully complete the test if they relied adequately on their memory and attention abilities. Also, an objective of the study was presented as to investigate age-related differences on the presented test. Participants were asked to indicate their age before they started the post-test.

Each trial of the computational estimation task (in both the threat and the control groups) started with a 400-ms fixation cross displayed at the centre of the screen, followed by the first priming problem, another 400-ms fixation cross, and the second priming problem. Then, again a 400-ms fixation cross was followed by the target problem. The timing of each response began when the problem appeared on the screen and ended when the experimenter pressed the space bar of the keyboard, the latter event occurring as soon as possible after the participants' responses. Participants were asked to calculate out loud to ensure which strategy was used. On each problem, the experimenter recorded participants' responses and strategy choices.

After the target problem of each trial, a 400 ms fixation cross appeared followed by four letters. The letters were displayed until the participant responded, pressing on the "L" key of the AZERTY keyboard when the four letters were only consonants (e.g., *trlc*) or only vowels (e.g., *aeio*), and on the "S" key when the four letters included both consonants and vowels (e.g., *ubqi*). A blank screen was finally displayed for 1000 ms at the end of each trial and before the next trial started.

Results

Results are reported in three main parts. We examined the effects of ABST first on participants' performance and then on strategy use, for priming and target problems separately. Lastly, we examined individual differences in ABST effects and conducted moderation analyses to determine what moderates ABST effects on participants' performances. In all results, unless otherwise noted, differences are significant to at least $p<.05$.

Effects of Age-Based Stereotype Threat on Participants' Performance

Participants' performance (mean estimation latencies and absolute percent deviations) on priming and target problems (see means in Table 8) were analyzed by 2(Group: threat, control) x 2(Session: pre-test, post-test) mixed-design ANOVAs, with repeated measures on the last factor.

Strategy performance on priming problems.

Participants were overall faster in the post-test compared to the pre-test session (pre-test – post-test = 1249 ms; $F(1,46) = 83.75$, $MSe = 37{,}845{,}001$, $\eta^2_p = 0.57$). The Group x Session interaction was significant ($F(1,46) = 7.37$, $MSe = 3{,}300{,}342$, $\eta^2_p = 0.14$). This interaction resulted from larger differences in estimation latencies between the pre-test and post-test sessions in the threat group (pre-test – post-test = 1620 ms; $F(1,46) = 70.41$, $\eta^2_p = 0.60$) than in the control group (pre-test – post-test = 878 ms; $F(1,46) = 20.71$, $\eta^2_p = 0.31$).

Analyses of absolute percent deviations revealed that participants' estimates deviated more from correct products in the threat group than in the control group (threat – control = 3.4%; $F(1,46) = 78.93$, $MSe = 135$, $\eta^2_p = 0.63$), and in the post-test session than in the pre-test session (post-test – pre-test = 2.6%; $F(1,46) = 45.52$, $MSe = 156$, $\eta^2_p = 0.49$. The Group x Session interaction came out significant ($F(1,46) = 55.68$, $MSe = 191$, $\eta^2_p = 0.55$). Further contrasts revealed that the difference in percent deviations between the threat and control groups was significant in the post-test session (threat – control = 6.2%; $F(1,46) = 79.19$, $MSe = 459$, $\eta^2_p = 0.63$), but not in the pre-test session (threat – control = 0.5%; $F<3$). Comparing the pre-test and post-test sessions within each group showed that the difference in percent deviations between the two sessions was significant in the threat group (post-test – pre-test = 5.4%; $F(1,23) = 53.22$, $MSe = 347$, $\eta^2_p = 0.69$), but not in the control group (pre-test – post-test = 0.3%; $F<2$).

Table 8. Mean Percentages of Use of the Better Strategy, Mean Estimation Latencies (in ms), and Absolute Percent Deviations (in parentheses) on Priming and Target problems in the Threat and Control Groups.

Sessions	Threat	Control	*Means*
	Priming Problems		
	Percentages of Use of the Better Strategy (%)		
Pre-Test	96	95	95
Post-Test	68	97	82
Means	82	96	89
	Mean Estimation Latencies in ms (and percent deviations)		
Pre-Test	5988 (10.4)	5495 (9.9)	5741 (10.1)

Post-Test	4368 (15.8)	4616 (9.6)	4492 (12.7)
Means	5178 (13.1)	5055 (9.7)	5116 (11.4)
	Target Problems		
	Percentages of Use of the Better Strategy (%)		
Pre-Test	84	83	83
Post-Test	56	84	70
Means	70	83	76
	Mean Estimation Latencies in ms (and percent deviations)		
Pre-Test	6883 (16.5)	6133 (15.3)	6508 (15.9)
Post-Test	5417 (18.4)	5440 (15.4)	5428 (16.9)
Means	6150 (17.5)	5786 (15.4)	5968 (16.4)

Strategy performance on target problems.

Participants were overall faster in the post-test compared to the pre-test session (pre-test – post-test = 1080 ms; $F(1,46) = 40.47$, $MSe = 27{,}951{,}819$, $\eta^2_p = 0.47$). The Group x Session interaction ($F(1,46) = 5.17$, $MSe = 3{,}576{,}328$, $\eta^2_p = 0.1$) was significant. This interaction resulted from larger differences in estimation latencies between the pre-test and post-test sessions in the threat group (pre-test – post-test = 1465 ms; $F(1,46) = 37.29$, $\eta^2_p = 0.45$) than in the control group (pre-test – post-test = 693 ms; $F(1,46) = 8.34$, $\eta^2_p = 0.15$).

Analyses of absolute percent deviations revealed that participants' estimates deviated more from correct products in the threat group than in the control group (threat – control = 2.1%; $F(1,46) = 18.09$, $MSe = 50$, $\eta^2_p = 0.28$), and in the post-test than in the pre-test session (post-test – pre-test = 1.0%; $F(1,46) = 8.90$, $MSe = 24$, $\eta^2_p = 0.16$). The Group x Session interaction was significant ($F(1,46) = 7.21$, $MSe = 19$, $\eta^2_p = 0.14$). Difference in percent deviations between the threat and control groups was significant in the post-test session (threat – control = 3.0%; $F(1,46) = 33.90$, $MSe = 104$, $\eta^2_p = 0.42$), but not in the pre-test session (threat – control = 1.2%; $F<3.0$). Comparing the pre-test and post-test sessions within each group showed that difference in percent deviations between the two sessions was significant in the threat group (post-test – pre-test = 1.9%; $F(1,23) = 8.36$, $MSe = 87$, $\eta^2_p = 0.27$), but not in the control group (pre-test – post-test = 0.1%; $F<1$).

Effects of Age-Based Stereotype Threat on Participants' Strategy Selection.

Mean percentages of use of the better strategy (see means in Table 8) were analyzed by 2(Group: threat, control) x 2(Session: pre-test, post-test) mixed-design ANOVAs, with repeated measures on the last factor.

Strategy selection on priming problems.

Overall, older adults selected the better strategy more often in the control than in the threat group (control – threat = 14%; $F(1,46) = 57.03$, $MSe = 2{,}359$, $\eta^2_p = 0.55$), and in the pre-test compared to the post-test session (pre-test – post-test = 13%; $F(1,46) = 43.82$, $MSe = 4{,}017$, $\eta^2_p = 0.48$). The Group x Session interaction was significant, $F(1,46) = 53.64$, $MSe = 4{,}916$, $\eta^2_p = 0.54$. The difference in percentages of use of the better strategy between the control and the threat groups was significant in the post-test session (control – threat = 29%; $F(1,46) = 61.33$, $MSe = 9{,}633$, $\eta^2_p = 0.57$), but not in the pre-test session (threat – control = 1%; $F<1$). Comparing the pre-test and the post-test sessions within each group showed that the difference in percentages of use of the better strategy between the two sessions was significant in the threat group (pre-test – post-test = 28%; $F(1,23) = 51.96$, $MSe = 8{,}910$, $\eta^2_p = 0.69$) but not in the control group (post-test – pre-test = -2%; $F<2$).

Strategy selection on target problems.

Overall, older adults selected the better strategy more often in the control than in the threat group (control – threat = 13%; $F(1,46) = 51.07$, $MSe = 2{,}193$, $\eta^2_p = 0.53$), and in the pre-test compared to the post-test session (pre-test – post-test = 13%; $F(1,46) = 55.13$, $MSe = 4{,}280$, $\eta^2_p = 0.54$). The Group x Session interaction was significant, $F(1,46) = 65.18$, $MSe = 5{,}060$, $\eta^2_p = 0.59$. The difference in percentages of use of the better strategy between the control and the threat groups was significant in the post-test session (control – threat = 28%; $F(1,46) = 94.47$, $MSe = 9{,}436$, $\eta^2_p = 0.67$), but not in the pre-test session (threat – control = 1%; $F<1$). Comparing the pre-test and the post-test sessions within each group showed that the difference in percentages of use of the better strategy between the two sessions was significant in the threat group (pre-test – post-test = 28%; $F(1,23) = 121.41$, $MSe = 9{,}324$, $\eta^2_p = 0.84$) but not in the control group (post-test – pre-test = 1%; $F<1$).

Strategy repetitions.

We conducted a 2(Group: threat, control) x 2(Session: pre-test, post-test) x 2(Trial: repeated/unrepeated) mixed design ANOVA on mean percentages of strategy repetitions, with repeated measures on the last two factors (see means in Table 9). Strategy repetition was coded 1 if participants used the same strategy on the two priming problems and target problems; otherwise, it was coded 0.

Table 9. Mean Percentages of Strategy Repetitions on Repeated Trials (priming and target problems were of the same type) and Unrepeated Trials (priming and target problems were of the same type) on Priming and Target problems in the Threat and Control Groups.

Session	Threat	Control	*Means*
	Repeated Trials		
Pre-Test	90	87	88
Post-Test	68	91	80
Means	79	89	84
	Unrepeated Trials		
Pre-Test	22	22	22

Post-Test	56	23	40
Means	39	23	31

Participants repeated the same strategy across two successive problems more often on repeated trials than on unrepeated trials (repeated – unrepeated = 53%; *F*(1,46) = 810.86, *MS*e = 137,067, η^2_p = 0.94). Most interestingly, the Group x Session x Trial interaction was significant (*F*(1,46) = 65.37, *MS*e = 10,150, η^2_p = 0.58). During post-test, participants repeated the same strategy on priming and target problems more often in the control than in the threat group for repeated trials (control - threat = 23.7%; *F*(1,46) = 31.31, *MS*e = 6,745, η^2_p = 0.41) and less often in the control than in the threat condition for unrepeated trials (control - threat = -32.3%; *F*(1,46) = 43.95, *MS*e = 12,545, η^2_p = 0.49). No differences were found between threat and control groups in the pre-test session for either repeated or unrepeated trials (*F*s<1.0). Comparing the pre-test and the post-test sessions within each group showed that the difference in percentages of strategy repetitions between the two sessions was significant only in the threat group for repeated trials (pre-test – post-test = 22.3%; *F*(1,23) = 17, η^2_p = 0.42) and unrepeated trials (pre-test – post-test = 33.2%; *F*(1,23) = 70, η^2_p = 0.75). No significant differences in percentages of strategy repetitions were found between the two sessions in the control group (*F*s<2.0).

Moderation Analyses

To determine whether general (processing speed) and/or specific (arithmetic fluency) processing resources moderated threat effects in older adults, we first conducted a series of independent sample *t*-tests to ensure that our threat and control groups did not differ in performance on our moderator variables. There were no differences between the two groups for any of these variables, all *p*s >.7.

We then conducted moderation analyses, to test the moderation effects of arithmetic fluency and processing speed on our dependent variable, absolute percent deviations. Using

PROCESS (5000 bootstrapped resamples; Model 2; Hayes, 2013), absolute percent deviations were regressed on Group (dummy coded: 1 = Control and -1 = Threat) with *z*-scores of arithmetic fluency entered as our first moderator variable and *z*-scores of processing speed entered as our second moderator variable (see main effects and interactions in Table 10).

Table 10. Linear models of predictors of threat effects (post-test – pre-test).

Predictors	*Coeff (b)*	*SE*	*t*	*p*	*Confidence Intervals*
Model 2 (Arithmetic Fluency + Processing Speed on Percent Deviations)					
Constant	1.959	.206	9.523	.000	[1.544, 2.374]
Arithmetic Fluency (AF)	-0.948	.270	-3.512	.001	[-1.494, -0.404]
Processing Speed (PS)	-0.195	.254	-0.767	.446	[-0.707, 0.317]
Group	-1.823	.206	-8.864	.000	[-2.238, -1.408]
Group x AF	1.114	.270	4.122	.000	[0.568, 1.658]
Group x PS	0.158	.254	.623	.536	[-0.354, 0.671]

Results showed that our moderation Model 2 with arithmetic fluency and processing speed as moderators was overall significant ($F(5,42) = 28.47$, $MSe = 1$, $R^2_{\text{overall effect}} = 0.79$, $p < .001$). While the two Group x Arithmetic Fluency and Group x Processing Speed interactions together accounted for 18% of the variance in performance ($F(2,42) = 18.79$; $\Delta R^2 = .185$; $p < .001$), the moderation by arithmetic fluency interaction uniquely accounted for 14% of the variance ($F(1,42) = 16.99$; $\Delta R^2 = .142$; $p < .001$), whereas the moderation by processing speed interaction accounted for only 0.2% of the variance (i.e., the interaction Group x Processing Speed[3] was not significant while controlling for Threat x Arithmetic Fluency; $F(1,42) = 0.38$;

[3] While our findings suggest that processing speed does not moderate threat effects in arithmetic domain, we found a significant bivariate relation between processing speed and percent deviations ($r = -.431$). We therefore conducted a simple moderation analysis (Model 1) and found a significant Group x Processing speed interaction suggesting that the relation between threat and percent deviations was moderated by processing speed ($F(1,44) = 5.78$;

ΔR^2= .002; p = .536). Thus, we found that arithmetic fluency was the best moderator. To interpret this moderation effect, we further examined the simple slopes (see Figure 4), that is the conditional effects of threat on absolute percent deviations evaluated at 1 SD below and above arithmetic fluency scores. of arithmetic fluency scores. We found that with increases in arithmetic fluency the strength of the relation between threat and performance goes from a large negative effect (b = -2.93, 95%CI [-3.64, -2.21], SE = 0.355, t = -8.25, p < .001) to a smaller negative effect (b = -0.718, 95%CI [-1.36, -0.07], SE = 0.320, t = -2.24, p =.030).

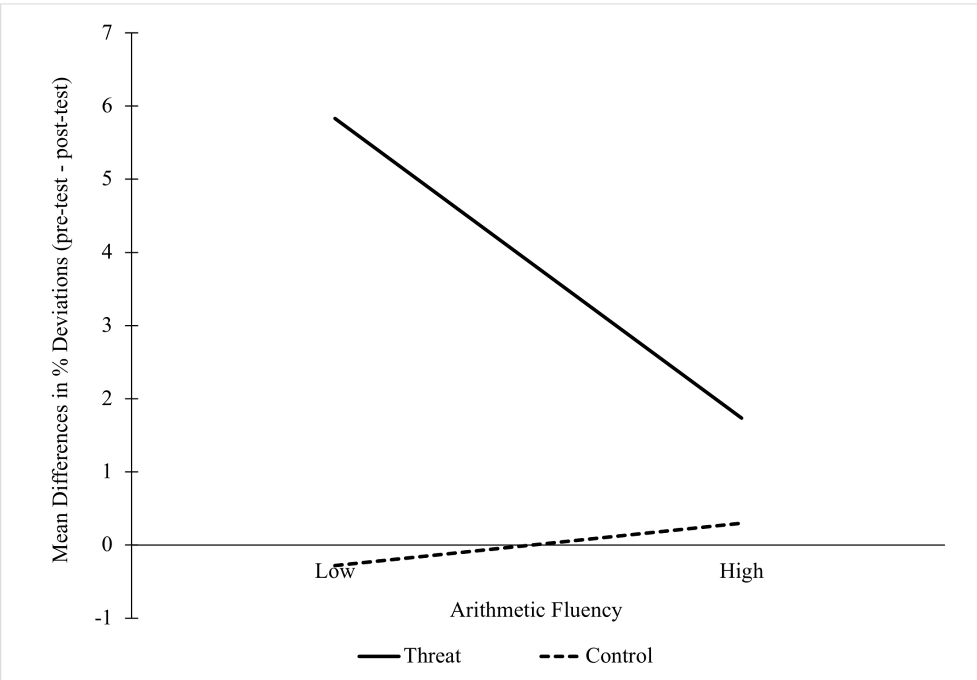


Figure 4. *Relations between Threat and Performance as a function of Low (-1 SD) and High (+1 SD) levels of Arithmetic Fluency.*

General Discussion

The three experiments reported here reflected the convergence of several unique

ΔR^2= .049; p = .020). It would be interesting for future studies to investigate processing speed as a moderator of threat effects in domains other than arithmetic.

features. We investigated ABST effects in arithmetic, a domain where ABST effects have never been tested and where age-related differences in performance are smaller than in many cognitive domains. We used both between (Expts. 1 and 2) and within-participants (Expt. 3) designs. We investigated strategic aspects of participants' performance under threat and control conditions both directly and indirectly. Participants accomplished an arithmetic problem verification task (where strategies are inferred from variations in participants' performance as a function of stimulus sets) in Experiment 1 or computational estimation tasks (where external behavioral evidence of strategies can be collected independently of participants' performance) in Experiments 2 and 3. Moreover, pre-test/post-test design of Experiment 3, combined with measures of individuals' general and specific processing resources enabled us to investigate individual differences in ABST and to test the role of different cognitive moderators. Our results establish that ABST effects are independent of the cognitive domain or tasks in which they are studied and of pre-experimental differences in young and older adults' baseline performance. They provide converging evidence that ABST effects occur via changes in which strategies are used and how strategies are executed. They also document the role of several moderators on different levels of individuals' sensitivity to ABST effects. These findings have important implications to further our understanding of ABST effects as well as effects of other stereotype threats, and, more generally, of how psychosocial factors exert a crucial role on age-related changes in human cognition. We next discuss these implications.

When does ABST impair older adults' cognitive performance?

A number of previous studies established that older adults' performance decreases under stereotype threat (see Barber, 2017; Lamont et al., 2015, for reviews). This has been found in such a wide variety of domains as driving, employment outcomes, cognitive, and physical performance. An important limitation of previous studies on ABST is that they have been mostly conducted in domains and tasks where pronounced age-related differences are well established. In contrast, this study is the first to report ABST effects in a domain (arithmetic) and in two types of tasks (problem verification and computational estimation) where there are no or reduced age-related declines (see Duverne & Lemaire, 2005; Uittenhove & Lemaire, 2015, 2018, for reviews). By observing threat effects on older adults' arithmetic performance, the present findings importantly demonstrate that ABST effects exist independently of older adults' baseline performance and independently of age-related declines in cognitive performance.

Concerns about whether stereotype threat effects occur independently of baseline performance have been discussed in the broader literature on stereotype threat (see Jussim, Crawford, & Rubinstein, 2015; Wax, 2009 for reviews). Such concerns led some researchers to run statistical analyses that include baseline performance as a covariate, when the stigmatized and non-stigmatized groups differ in baseline performance (see Jussim, Jarret, Stephanie, Sean & Jose, 2016 for a review). Although effects of stereotype threat have been found when baseline performance was used as a covariate, this analytical approach raised doubts about the power of stereotype threat and has been criticized (e.g., Sackett, Hardison, & Cullen, 2004). These criticisms refer to the misinterpretation that equal adjusted means in ANCOVA are equivalent to equal means. The failure to recognize the implications of the statistical adjustments for existing prior differences between the control and threat groups has often led to incorrect conclusions in the stereotype threat literature that the differences between the two groups disappear when stereotype threat is removed (see Wicherts, 2005, for a discussion). The potential impact of baseline performance is no longer a concern when, like here, participants are tested on tasks and in domains with no (or much reduced) age-related differences in baseline performances. Here, consistent with previous findings (Hinault et al., 2015a, b), Experiment 1 showed no group differences in young and older adults tested under the control condition. This ensures that observed ABST effects were independent of older adults' baseline performance.

Future studies can adopt such a method to avoid skeptical analyses, demonstrate the strength of ABST effects, discover their conditions of occurrence and their underlying mechanisms.

Our findings may not be limited to arithmetic. Deleterious effects of age-based stereotype threat may also occur in other domains where we often see much less age-related differences, like moral reasoning (e.g., Margoni, Geipel, Hadjichristidis, & Surian, 2018; McNair, Okan, & Bruin, 2018), social wisdom (e.g., Grossmann et al., 2010), decision making (e.g., see Löckenhoff, 2018, for a review). Most importantly, they may also occur in domains where ABST effects have been less investigated (e.g., sensory acuity, language processing).

How does stereotype threat impairs older adults' performance?

We found that older adults recruited different sets of mechanisms (or strategies) to accomplish the same target task under threat and control groups. Moreover, they differed in how they executed the same mechanisms (or strategies) when they did not use different mechanisms. We found these results on two types of arithmetic tasks that examined changes in strategies either indirectly (arithmetic problem verification task in Expt. 1) or directly (computational estimation tasks in Expts. 2 and 3). These strategy findings have important implications to further our understanding of mechanisms underlying ABST effects.

That older adults used the better strategy less often and executed available strategies less efficiently under threat here are consistent with previous proposals or empirical findings (e.g., Hess et al., 2003; Lemaire et al., 2018). Indeed, like Hess et al. who found some indirect evidence that threat reduced the use of more efficient memory strategies, we also inferred from our data (Exp. 1) that threat disrupted the use of more efficient arithmetic strategies. For example, longer overall verification latencies under threat can be explained as a result of older adults selecting the exhaustive calculation strategy to verify a false equation, when a quick plausibility checking strategy was used in the control condition. This might seem peculiar at first as to why older adults under threat may choose a more demanding exhaustive calculation strategy over a quick plausibility checking strategy, however we can explain it in two ways. First the use of quick plausibility checking strategy requires attention to semantic features of the problems (Krueger, 1986; Krueger & Hallford, 1984), that is cognitive flexibility involving a high level of attentional control. It is very much likely that threat affected older adults' cognitive flexibility disabling them to pay attention to such semantic details of the target arithmetic problems, hence they use the more resource demanding strategy over easy plausibility checking strategy. Second most importantly, it might have been easier for older adults to use one strategy that can be applied to all problems (i.e., both true and false problems) to avoid switching between strategies which could be resource demanding, especially under threat. Since the quick plausibility check could not be applied to all problems, older adults under threat used the more resource demanding strategy on all problems.

However, Experiment 1 was limited in making it impossible to have external behavioral evidence of such strategy disruption, because we could not test directly whether ABST led older adults to use the more efficient plausibility checking strategy less often or to execute it more poorly. However, with a Choice/No-Choice method, we found that threat occurred via disrupted strategy use and execution in older adults in Experiment 2. Moreover, our mediation analyses tested the hypothesis that strategy use mediates threat effects in older adults. We found a significant attenuation of the threat-related variance in performance when strategy use was statistically controlled (Expts. 2 & 3). We also found reduced or exacerbated threat effects on different measures of complex-arithmetic performances (Exps. 2 & 3). One distinct pattern was a relatively rapid performance by older adults under threat, often as rapid as those in control group (e.g., heterogeneous problems in Exp. 2 & target problems in Exp. 3), and sometimes faster than those in control group (e.g., homogeneous problems in Exp. 2 & priming problems in Exp.3) but accompanied by a dramatically higher rate of deviation from the correct answer. This can be explained again as a result of older adults being less careful under threat at selecting the better strategy when given a choice that leads them to be quicker in responding while the responses were being compromised.

An additional important finding in our study concerns the differences in strategy repetitions under threat and control groups. We found increased strategy repetitions under threat (Expts. 2 & 3), especially on trials where it was inappropriate to repeat strategies to select the better strategy. Interestingly, increased rates of inappropriate strategy repetitions under threat suggests that ABST led older adults to be less efficient to inhibit the just executed strategy, to re-activate the two available strategies in working-memory while encoding the operands, to analyze problem features (i.e., size of unit digits) crucial for better strategy selection, and to choose the better strategy. This is consistent with previous findings that the ability to select a strategy requires inhibitory control to suppress competing strategies and cognitive flexibility to alternatively switch between strategies (e.g., Ardiale, Hodzik, & Lemaire, 2012; Ardiale & Lemaire, 2012; Hodzik & Lemaire, 2011; Lemaire and Lecacheur, 2010; Luwel, Onghena, Torbeyns, Schillemans, & Verschaffel, 2009). This is also consistent with the hypothesis that stereotype threat impairs executive control mechanisms (Croizet, Després, Gauzins, Huguet, & Leyens, 2004; Mazerolle et al., 2012; Schmader & Johns, 2003; Schmader et al., 2008; Schmader & Beilock 2012).

Another finding consistent with the hypothesis that ABST decreased older adults' executive control processes concerns participants' performance on trials where strategy selection was controlled (Expt. 2, No-Choice condition). We found larger poorer strategy effects (i.e., differences in participants' performance while executing the poorer and the better strategies) in the threat group than in the control group. This was mostly the result of increased latencies under threat condition when participants were cued to execute the poorer strategy. No differences were found between control and threat groups when participants were required to execute the better strategy. Previous studies of poorer strategy effects revealed that participants first activate the better strategy when they encode problems. If the cued strategy is the poorer strategy, participants then inhibit the more or less automatically activated better strategy before executing the required, poorer strategy (e.g., Hinault et al., 2014, 2016; Lemaire et al., 2004; Lemaire & Hinault 2014; Lemaire & Lecacheur, 2010). Here, ABST likely impaired older adults' executive resources used to inhibit the most automatically activated, better strategy and to focus on the execution of the cued poorer strategy.

Although the present findings are consistent with the hypothesis that ABST impaired older adults' executive resources, it is also plausible that other mechanisms contributed to how ABST impacts participants' performance via strategy selection and execution. Indeed, several mechanisms responsible for how stereotype threat results in poorer performance have been hypothesized in the stereotype threat literature. And it is possible that these mechanisms, or some of them, underlie the present findings. For example, ABST might impact adaptive strategy selection by lowering participants' performance expectations (e.g., Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti, 2003; Desrichards & Kopetz, 2005) which may lower the amount of effort exerted to make adaptive strategy choices. Similarly, increased anxiety/arousal (e.g., Abrams et al., 2006, 2008) might affect cognitive processes controlling strategy use. More generally, the so-called older adults' regulatory focus (e.g., Barber & Mather, 2014; Higgins, 1997, 1999), or how older adults approach cognitive tasks, may determine which strategies are deployed and how these deployed strategies are implemented while participant accomplish cognitive tasks.

It is important to note that a strategy approach of stereotype threat effects is not inconsistent with other proposals in the literature like the integrated process model (e.g., Schmader, et al., 2008) or the regulatory focus model (e.g., Barber & Mather, 2013). The regulatory focus model of stereotype threat argues that people respond to threat by becoming more prevention-focused, that is by becoming more vigilant to avoid losses and negative outcomes (see, Barber, 2017; Popham & Hess, 2015, 2016; Seibt & Forster, 2004). It is likely that in our experiments ABST led older adults to adopt a more prevention focus (i.e. increased vigilance towards avoiding errors), leading older adults under threat to repeat strategies on consecutive trials more often, given that strategy repetitions helped them to execute the repeated strategy more quickly, with fewer errors, and with fewer resources. In other words, older adults tended to avoid errors and minimized estimation latencies by not switching between strategies that they were not sure they could execute as efficiently. An alternative, though not contradictory, account is the integrated process model that suggests that threat triggers negative affective responses and increases task monitoring. As these processes are cognitively effortful, this reduces the amount of executive resources available to perform the task, resulting in impaired performance. ABST effects as found in our experiments on older adults' strategy use and execution can easily be accounted for by assuming that threat taxes executive resources, as strategy use and execution in arithmetic are known to be highly demanding on executive control processes (e.g., Ardiale et al., 2012; Beilock & DeCaro, 2007; Hodzik & Lemaire, 2011; Imbo, Vandierendonck, & Vergauwe, 2007; Lemaire & Lecacheur, 2010).

How individual differences alter ABST experience?

Previous studies found that some older adults' performance decreased more than others' under ABST conditions. Indeed, larger ABST effects were found in older individuals who tend to be younger, more highly educated, who tend to have more negative self-perceptions of aging, who have experienced or imagined low quality intergenerational interactions (e.g., Abrams, 2006; Abrams et al., 2008), or higher levels of dementia-worry (e.g., Fernández-Ballesteros et al., 2015; Fresson et al., 2017; Hess et al., 2009). The present study complements previous studies by investigating the magnitudes of ABST effects in each individual with a pretest-posttest design. This made it possible to quantify the amount of interference of ABST on participants' performance and to determine to what extent some older individuals were able to resist more efficiently to ABST effects. Second, we assessed both domain-general (processing speed) and domain-specific (arithmetic fluency) processing resources and determined which resources moderate individual differences in ABST effects. Our findings deepen our understanding of how older individuals vary in their sensitivity to ABST effects.

We found that basic arithmetic skills interacted with ABST to modulate participants' arithmetic performance. Older adults with higher arithmetic fluency showed reduced ABST effects relative to older adults with lower arithmetic fluency. These findings can be explained by assuming that if ABST limits processing resources under threat, individuals with higher processing resources to start with will probably deplete their limited processing resources under threat much less than those who start with lower processing resources. As a consequence, individuals with higher processing resources will have much more processing resources left free for the target task than individuals with lower processing resources. Such higher level of available processing resources will enable individuals to perform the task better than individuals with lower processing resources. Our findings are consistent with previous results that showed close relations between processing resources and stereotype threat. For example, Régner et al. (2010) found that women with higher working-memory capacity resisted stereotype threat effects more effectively relative to those with lower working-memory capacity (see also Beilock et al., 2007; Beilock & Carr, 2005; Inzlicht, McKay, & Aronson, 2006; Rydell, McConnell, & Beilock, 2007).

Combined together, findings from research on stereotype threat suggest that whichever stigmatized group, individuals with higher cognitive resources resist (or are less vulnerable) to stereotype threat. Here, in the context of arithmetic, domain-specific resources (arithmetic fluency) helped some older individuals to be less impaired by ABST. It is possible that individuals with higher processing resources were especially good at focusing their attention on task-relevant properties, and at ignoring threat related worries that likely perturbed those with lower cognitive resources. This should not lead to deny the role of domain-general processes on ABST effects. As influences of domain-general processes, like working-memory capacity has already been found (e.g., Régner et al., 2010), it is possible that involvement of domain-general and domain-specific processes may depend on the task or domains in which ABST effects occur. Speculatively, it could be that domain-specific resources are most heavily involved when participants' performance depends on expertise in the domain where ABST effects are assessed (like arithmetic) and that domain-general resources are most important in domain-general cognitive functions (like memory, attention, abstract inferences).

At a general level, future studies on ABST, and other stereotype threats, would make important progress by taking individual differences into account. Here, we assessed individual differences in (domain-general and domain-specific) cognitive resources. Other psychological factors have been assumed to modulate effects of stereotype threat (see Barber, 2017, for a recent review). Such psychological factors include affective factors (e.g., anxiety, evaluation apprehension, or stress) and motivational factors (e.g., decreased performance expectations, increased avoidance goals, or changes in one's motivation to do well). These factors may also contribute to individual differences in ABST.

Implications for understanding the role of psycho-social factors in cognitive aging.

The strategy perspective adopted in this study enabled us to do more than simply observing that older participants underperform in stereotype threat conditions. It afforded us to characterize ABST effects more precisely, specifying their nature, the conditions of their occurrence, and the underlying mechanisms. In the current study, we found that stereotype threat always occurred via disrupted strategy use and execution in older adults. This suggests that examining strategic variations is a promising perspective to more precisely describe and provide a mechanistic account of ABST effects. It also suggests that such a strategy perspective could be adopted to further investigate effects of any situational characteristics that have been found to modulate age-related changes in human cognition. Such situational characteristics involve, for examples, social connectedness (Amieva et al., 2010; Crooks, Lubben, Petitti, Little, & Chiu, 2008; Fratiglioni, Paillard-Borg, & Winblad, 2004), social support (Béland et al., 2005; Seeman, Lusignolo, Albert, & Berkman, 2001), or control beliefs (Bielak et al., 2007; Lachman & Andreoletti, 2006). Above and beyond cognitive resources and situations, age-related changes in cognition have been found to interact with a number of other factors, like well-being (Allerhand, Gale, & Deary, 2014; Gerstorf, Lövdén, Röcke, Smith, & Lindenberger, 2007), negative affect or depression (Ownby, Crocco, Acevedo, John, & Loewenstein, 2006; Royall, Palmer, Chiodo, & Polk, 2012), subjective feelings of loneliness (Holwerda et al., 2014), individual personality traits (Caselli et al., 2016; Kuzma, Sattler, Toro, Schonknecht, & Schroder, 2011; Luchetti, Terracciano, Stephan, & Sutin, 2015; Matthews, 2009), emotions (Moore & Oaksford, 2002; Pessoa, 2012), or physical activity (Bherer, Erickson, & Liu-Ambrose, 2013). To understand how these factors modulate cognitive aging, a strategy perspective enables to describe how these factors change the mechanisms that older adults adopt to accomplish cognitive tasks in different contexts (see Lemaire, 2016, for a deeper discussion on strategy perspective). A strategy approach has already proved fruitful to provide a better mechanistic account of the effects of several modulators of cognitive aging (e.g., see Bjälkebring, Västfjäll, & Johansson, 2013, for emotions; Lachman & Andreoletti, 2006, for control beliefs). As all cognitive tasks involve a sequence of mental procedures (or strategies), it is highly likely that adopting a strategy perspective to investigate cognitive aging would help us to better understand how and to what extent different psychosocial factors influence age-related changes in human cognition.

Context

The two senior authors on this paper are in the same research institution. They had plenty of opportunities to exchange about their work and research agenda, PL on strategic aspects of mathematical cognition and on cognitive aging and IR on effects of stereotype threat. When PL and PN heard IR discussing her work on effects of stereotype threat, they raised a number of issues (e.g., how do we know if mechanisms used by participants to accomplish tasks under threat and nonthreat conditions are the same or different? What if these mechanisms were qualitatively different?). When IR and PN read and heard PL's previous works on strategic aspects of cognition and cognitive aging, they thought that his conceptual and methodological approaches were relevant to study effects of stereotype threat. Discussions in Lab meetings convinced them to join their efforts to investigate effects of stereotype threat (and ABST in particular) with a new (strategy) perspective that proved useful in past research to investigate mechanisms underlying mathematical performance as well as cognitive aging. Given their expertise, they decided to conduct the present three experiments on ABST effects in arithmetic, a domain where ABST had never been investigated before.

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

Abrams, D., Eller, A., & Bryant, J. (2006). An age apart: The effects of intergenerational contact and stereotype threat on performance and intergroup bias. *Psychology and Aging, 21(4)*, 691–702. https://doi.org/10.1037/0882-7974.21.4.691

Allerhand, M., Gale, C., Deary, I. (2014). The dynamic relationship between cognitive function and positivewell-being in older people: A prospective study using the English Longitudinal Study of Aging. *Psychol. Aging, 29(2),* 306–318. https://doi.org/10.1037/a0036551

Anders, R., Hinault, T., & Lemaire, P. (2018). Heuristics versus direct calculation, and age-related differences in multiplication: an evidence accumulation account of plausibility decisions in arithmetic. *Journal of Cognitive Psychology, 30(1),* 18–34. https://doi.org/10.1080/20445911.2017.1397676

Amieva, H., Stoykova, R., Matharan, F., Helmer, C., Antonucco, T., C., Dartigues, J., F. (2010). What aspects of social network are protective for dementia? Not quantity but the quality of social interactions is protective up to 15 years later. *Psychosomatic Medicine, 72(9),* 905–911. https://doi.org/10.1097/psy.0b013e3181f5e121

Ardiale, E., Hodzik, S., & Lemaire, P. (2012). Aging and Strategy Switch Costs: A study in Arithmetic Problem Solving. *L'Année Psychologique, 112,* 345–360. https://doi.org/10.4074/s0003503312003028

Ardiale, E., & Lemaire, P. (2012). Age-related differences in within-item strategy switching. *Psychology & Aging*, 27, 1138–1151. https://doi.org/10.1037/a0027772

Armstrong, B., Gallant, S., N., Li, L., Patel, K., & Wong, B. (2017). Stereotype threat Effects on older adults' episodic and working memory: A meta-analysis. *The Gerontologist. 57(S2*), S193–S205. https://doi.org/10.1093/geront/gnx056

Baltes, P. B., & Lindenberger, U. (1997). Emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging? *Psychology and Aging, 12*, 12–21. http://dx.doi.org/10.1037/0882-7974.12.1.12

Barber, S., J., & Mather, M. (2013a). Stereotype threat can both enhance and impair older adults' memory. *Psychological science, 24(12)*, 2522–2529. https://doi.org/10.1177/0956797613497023

Barber, S., J., Mather, M. (2013b). Stereotype threat can reduce older adults' memory errors. *The Quarterly Journal of Experimental Psychology, 66(10),* 1888–1895.

https://doi.org/10.1080/17470218.2013.840656

Barber, S., J., & Mather, M. (2014). Stereotype threat in older adults: When and why does it occur, and who is most affected? In P. Verhaeghen, & C. Hertzog (Eds.), *The Oxford handbook of emotion, social cognition, and problem solving during adulthood* (pp. 302–320). Oxford, UK: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199899463.001.0001

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., Mather, M., & Gatz, M. (2015). How stereotype threat affects healthy older adults' performance on clinical assessments of cognitive decline: The key role of regulatory fit. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 70(6)*, 891–900. https://doi.org/10.1093/geronb/gbv009

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., & Carr, T., H. (2005). When High-Powered People Fail. *Psychological Science, 16(2),* 101–105. https://doi.org/10.1111/j.0956-7976.2005.00789.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

Beilock, S., L., Rydell, R., J., & McConnell, A., R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spill over. *Journal of Experimental Psychology: General, 136(2)*, 256–276. https://doi.org/10.1037/e633982013-073

Béland, F., Zunzunegui, M., V., Alvarado, B., Otero, A., Del Ser, T. (2005). Trajectories of cognitive decline and social relations. *J. Gerontol. B Psychol. Sci. Soc. Sci, 60(6),* 320–330. https://doi.org/10.1093/geronb/60.6.p320

Bielak, A., A., M., Hultsch, D., F., Levy-Ajzenkopf, J., MacDonald, S., W., S., Hunter, M., A., Strauss, E. (2007). Short-term changes in general and memory-specific control beliefs and their relationship to cognition in younger and older adults. *International Journal of Aging and Human Development*, *65(1),* 53–71. https://doi.org/10.2190/g458-x101-0338-746x

Bjälkebring, P., Västfjäll, D., Johansson, B. (2013). Regulation of experienced and anticipated regret for daily decisions in younger and older adults in a Swedish one-week diary study. *GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry, 26(4),* 233–241. https://doi.org/10.1024/1662-9647/a000102

Bherer, L., Erickson, K., I., & Liu-Ambrose, T. (2013). A review of the effects of physical activity and exercise on cognitive and brain functions in older adults. *Journal of Aging Research,* 657508. https://doi.org/10.1155/2013/657508

Brubaker, M., S., & Naveh-Benjamin, M. (2018). The effects of stereotype threat on the associative memory deficit of older adults. *Psychology and Aging, 33(1),* 17–29. https://doi.org/10.1037/pag0000194

Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. *European Journal of Social Psychology, 33(2),* 267–285. https://doi.org/10.1002/ejsp.145

Caselli, R., Dueck, A., Locke, D., Henslin, B., Johnson, T., Woodruff, B., Hoffman-Snyder, C., Geda, Y. (2016). Impact of Personality on Cognitive Aging: A Prospective Cohort Study. *Journal of the International Neuropsychological Society, 22(7),* 765–776. https://doi.org/10.1017/s1355617716000527

Chaparro, A., & Alton, J. (2000). Age-related differences in driving performance and target identification. *Proceedings of the Human Factors and Ergonomics Society, 44(24),* 456-459. https://doi.org/10.1177/154193120004402416

Chasteen, A., L., Bhattacharyya, S., Horhota, M., Tam, R., & Hasher, L. (2005). How feelings of stereotype threat influence older adults' memory performance. *Experimental Aging Research, 31(3*), 235–260. https://doi.org/10.1080/03610730590948177

Chasteen, A., L., Kang, S., K., & Remedios, J., D. (2012). Aging and stereotype threat: Development, process, and interventions. In M. Inzlicht & T. Schmader (Eds.), *Stereotype threat: Theory, process, and application* (pp. 202–216). New York, NY, US: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199732449.003.0013

Craik, F., I., M., and Byrd, M. (1982) Aging and cognitive deficits: The role of attentional resources. In F. I. M. Craik & S. E. Trehub (Eds.), *Aging and Cognitive Processes,* (pp. 191–211). New York: Plenum Press. https://doi.org/10.1007/978-1-4684-4178-9_11

Croizet, J., C., Després, G., Gauzins, M., Huguet, P., Leyens, J. (2004). Stereotype threat undermines performance by triggering a disruptive mental load. *Personality and Social Psychology Bulletin, 30 (6),* 721–731. https://doi.org/10.1177/0146167204263961

Crooks, V., Lubben, J., Petitti, D., Little, D., Chiu, V. (2008). Social network, cognitive function, and dementia incidence among elderly women. *American Journal of Public Health, 98(7),* 1221–1227. https://doi.org/10.2105/ajph.2007.115923

Deltour, J., J. (1993). Echelle de Vocabulaire de Mill Hill de J.C. Raven. Adaptation Française et Normes Européennes du Mill Hill et du Standard Progressive Matrices de Raven (PM38) [Mill Hill Vocabulary Scale. French Adaptation and European Standardization]. Braine-le-Château: Editions l'application des techniques modernes. https://doi.org/10.1007/978-1-4899-2151-2_8

Desrichard, O., & Kopetz, C. (2005). A threat in the elder: The impact of task instructions, self- efficacy and performance expectations on memory performance in the elderly. *European Journal of Social Psychology, 35(4)*, 537–552. https://doi.org/10.1002/ejsp.249

Duverne, S., & Lemaire, P. (2005). Aging and Arithmetic. In J. I. D. Campbell (Ed.), *The handbook of mathematical cognition,* (pp. 397–412). New York : Psychology Press.

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*, 175–191. https://doi.org/10.3758/bf03193146

Fernández-Ballesteros, R., Bustillos, A., Huici, C. (2015). Positive Perception of Aging and Performance in a Memory Task: Compensating for Stereotype Threat? *Experimental Aging Research, 41(4),* 410–425. https://doi.org/10.1080/0361073x.2015.1053757

Folstein, M., F., Folstein, S., E., McHugh, P., R. (1975). Mini-Mental State: a pratical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12(3),* 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

French, J., W., Ekstrom, R., B., Price, I., A. (1963). *Kit of Reference Tests for Cognitive Factors*. Princeton, NJ: Educational Testing Service. https://doi.org/10.21236/ad0410915

Fresson, M., Dardenne, B., Geurten, M., & Meulemans, T. (2017). The effect of stereotype threat on older people's clinical cognitive outcomes: Investigating the moderating role of dementia worry. *The Clinical Neuropsychologist. 31 (8),* 1–23. https://doi.org/10.1080/13854046.2017.1307456

Foos, P., W. (1989). Adult age differences in working memory. *Psychology and Aging, 4(3)*, 269–275. https://doi.org/10.1037//0882-7974.4.3.269

Fratiglioni, L., Paillard-Borg, S., Winblad, B. (2004). An active and socially integrated lifestyle in late life might protect against dementia. *The Lancet Neurology, 3(6),* 343–353.

https://doi.org/10.1016/s1474-4422(04)00767-7

Gaillard, M., Desmette, D., & Keller, J. (2011). Regulatory focus moderates the influence of age-related stereotypic expectancies on older adults' test performance and threat-based concerns. *European Review of Applied Psychology, 61(1)*, 23–29. https://doi.org/10.1016/j.erap.2010.11.001

Gerstorf, D., Lövdén, M., Röcke, C., Smith, J., Lindenberger, U. (2007). Well-being affects changes in perceptual speed in advanced old age: Longitudinal evidence for a dynamic link. *Developmental Psychology, 43(3),* 705–718. https://doi.org/10.1037/0012-1649.43.3.705

Gilmore, C., Göbel, S., Inglis, M. (2018). *An Introduction to Mathematical Cognition.* London: Routledge. https://doi.org/10.4324/9781315684758

Grimm, L., R., Markman, A., B., Maddox, W., T., & Baldwin, G., C. (2009). Stereotype threat reinterpreted as a regulatory mismatch. *Journal of Personality and Social Psychology, 96(2)*, 288–304. https://doi.org/10.1037/a0013463

Grossmann, I., Na, J., Varnum, M., E., Park, D., C., Kitayama, S., Nisbett, R., E. (2010). Reasoning about social conflicts improves into old age. *Proceedings of the National Academy of Sciences, 107(16),* 7246–7250. https://doi.org/10.1073/pnas.1001715107

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

Hayes, A., F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis. A Regression-Based Approach*. New York, NY : Guilford Press.

Head, D., Raz, N., Gunning-Dixon, F., Williamson, A., & Acker, J., D. (2002). Age-related differences in the course of cognitive skill acquisition: The role of regional cortical shrinkage and cognitive resources. *Psychology and Aging, 17(1),* 72–84.

https://doi.org/10.1037//0882-7974.17.1.72

Hess, T., M., Auman, C., Colcombe, S., J., & Rahhal, T., A. (2003). The impact of stereotype threat on age differences in memory performance. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 58(1)*, 3–11. https://doi.org/10.1093/geronb/58.1.P3

Hess, T., M., & Hinson, J., T. (2006). Age-related variation in the influences of aging stereotypes on memory in adulthood. *Psychology and Aging, 21(3)*, 621–625. https://doi.org/10.1037/0882-7974.21.3.621

Hess, T., M., Hinson, J., T., & Hodges, E., A. (2009). Moderators of and mechanisms underlying stereotype threat effects on older adults' memory performance. *Experimental Aging Research, 35(2)*, 153–177. https://doi.org/10.1080/03610730802716413

Higgins, E., T. (1997). Beyond pleasure and pain. *American Psychologist, 52 (12),* 1280–1300. https://doi.org/10.1037//0003-066x.52.12.1280

Higgins, E., T. (1999). Promotion and prevention as a motivational duality: Implications for evaluative processes. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 503–525). New York: Guilford Press. https://doi.org/10.4135/9781412956253.n164

Hinault T., Badier J., M., Baillet, S., Lemaire, P. (2017). The sources of sequential modulations of control processes in arithmetic strategies: a MEG study. *Journal of Cognitive Neuroscience, 29(6)*, 1033–1043. https://doi.org/10.1162/jocn_a_01102

Hinault T., Dufau S., Lemaire P. (2014). Sequential modulations of poorer-strategy effects during strategy execution: an event-related potential study in arithmetic. *Brain and Cognition. 91,* 123–130. https://doi.org/10.1016/j.bandc.2014.09.001

Hinault T., Lemaire P., Phillips, N. (2016). Aging and sequential modulations of poorer strategy effects: an EEG study in arithmetic problem solving. *Brain Research. 1630,* 144–158. https://doi.org/10.1016/j.brainres.2015.10.057

Hinault T., Lemaire P., Touron, D. (2016). Aging effects in sequential modulations of poorer-strategy effects during execution of memory strategies. *Memory 25(2),* 176–186. https://doi.org/10.1080/09658211.2016.1146300

Hinault, T., & Lemaire, P. (2017). Aging, rule-violation checking strategies, and strategy combination: An EEG study in arithmetic. *International Journal of Psychophysiology, Elsevier, 120,* 23–32. https://doi.org/10.1016/j.ijpsycho.2017.07.003

Hinault, T., Dufau, S., & Lemaire, P. (2015a). Strategy combination in human cognition: A behavioral and ERP study in arithmetic. *Psychonomic Bulletin & Review, 22(1),* 190–199. https://doi.org/10.3758/s13423-014-0656-8

Hinault, T., Tiberghien, K., & Lemaire, P. (2015b). Age-related differences in plausibility-checking strategies. *Journals of Gerontology : Psychological Sciences and Social Sciences, Series B, 71(4)*, 613–621. https://doi.org/10.1093/geronb/gbu178

Hodzik, S., & Lemaire, P. (2011). Inhibition and shifting capacities mediate adults' age-related differences in strategy selection and repertoire. *Acta Psychologica, 137,* 335-344. https://doi.org/10.1016/j.actpsy.2011.04.002

Holwerda, T., J., Deeg, D., J., Beekman, A., T., VanTilburg, T., G., Stek, M., L., Jonker, C., Schoevers, R., A. (2012). Feelings of loneliness, but not social isolation, predict dementia onset: results from the Amsterdam Study of the Elderly (AMSTEL). *Journal of Neurology Neurosurgery & Psychiatry, 85(2),* 135–42. https://doi.org/10.1136/jnnp-2012-302755

Huguet, P., & **Régner, I.** (2007). Stereotype Threat Among School Girls in Quasi-Ordinary Classroom Circumstances. *Journal of Educational Psychology, 99,* 545–560. https://doi.org/10.1037/0022-0663.99.3.545

Huguet, P., & **Régner, I.** (2009). Counter-stereotypic beliefs in math do not protect school girls from stereotype threat. *Journal of Experimental Social Psychology, 45,* 1024–1027.https://doi.org/10.1016/j.jesp.2009.04.029

Hummert, M., L. (1999). A social cognitive perspective on age stereotypes. In T. M. Hess & F. Blanchard-Fields (Eds.), *Social cognition and aging* (pp. 175-195). New York: Academic Press. https://doi.org/10.1016/b978-012345260-3/50009-4

Imbo, I., Vandierendonck, A., & Vergauwe, E. (2007). The role of working memory in carrying and borrowing. *Psychological research, 71(4),* 467–83. https://doi.org/10.1007/s00426-006-0044-8

Inzlicht, M., & Ben-Zeev, T. (2000). A Threatening Intellectual Environment: Why Females Are Susceptible to Experiencing Problem-Solving Deficits in the Presence of Males. *Psychological Science*, *11(5),* 365–371. https://doi.org/10.1111/1467-9280.00272

Inzlicht, M., McKay, L., & Aronson, J. (2006). Stigma as ego depletion: How being the target of prejudice affects self-control. *Psychological Science, 17,* 262–269. https://doi.org/10.1111/j.1467-9280.2006.01695.x

Joanisse, M., Gagnon, S., & Voloaca, M. (2013). The impact of stereotype threat on the simulated driving performance of older drivers. *Accident Analysis & Prevention, 50,* 530–538. https://doi.org/10.1016/j.aap.2012.05.032

Jussim, L., Jarret, T., C., Stephanie M., A., Sean, T., S., Jose, L., D. (2016). Interpretations and methods: Towards a more effectively self-correcting social psychology. *Journal of Experimental Social Psychology,* *66*, 116–133. https://doi.org/10.1016/j.jesp.2015.10.003

Jussim, L., Crawford, J., T., & Rubinstein, R., S. (2015). Stereotype (in)accuracy in perceptions of groups and individuals. *Current Directions in Psychological Science, 24,* 490–497. https://doi.org/10.1177/0963721415605257

Kadosh, R., C., & Dowker, A. (2015). *The Oxford Handbook of Numerical Cognition.* UK: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642342.001.0001

Kang, S., K., & Chasteen, A., L. (2009). The moderating role of age-group identification and perceived threat on stereotype threat among older adults. *The International Journal of Aging & Human Development, 69(3)*, 201–220. https://doi.org/10.2190/ag.69.3.c

Kite, M., E., & Johnson, B., T. (1988). Attitudes toward older and younger adults: A meta-analysis. *Psychology and Aging, 3(3)*, 233–244. https://doi.org/10.1037//0882-7974.3.3.233

Kite, M., E., Stockdale, G., D., Whitley, B., E., & Johnson, B., T. (2005). Attitudes toward younger and older adults: An updated meta-analytic review. *Journal of Social Issues, 61,* 241–266. https://doi.org/10.1111/j.1540-4560.2005.00404.x

Krueger, L. E. (1986). Why 2 x 2 = 5 looks so wrong: On the odd-even rule in product verification. *Memory & Cognition, 14(2),* 141–149. https://doi.org/10.3758/bf03198374

Krueger, L. E., & Hallford, E. W. (1984). Why 2 + 2 = 5 looks so wrong: On the odd-even rule in sum verification. *Memory & Cognition, 12(2),* 171–180. https://doi.org/10.3758/bf03198431

Kuzma, E., Sattler, C., Toro, P., Schonknecht, P., Schroder, J. (2011). Premorbid personality traits and their course in mild cognitive impairment: Results from a prospective population-based study in Germany. *Dementia and Geriatric Cognitive Disorders, 32(3),* 171–177. https://doi.org/10.1159/000332082

Lachman, M., E., Andreoletti, C. (2006). Strategy use mediates the relationship between control beliefs and memory performance for middle-aged and older adults. *Journal of Gerontology: Psychological Sciences, 61(2),* 88–94. https://doi.org/10.1093/geronb/61.2.p88

Lambert, A., E., Watson, J., M., Stefanucci, J., K., Ward, N., Bakdash, J., Z., & Strayer, D., L. (2016). Stereotype Threat Impairs Older Adult Driving. *Applied Cognitive Psychology, 30(1),* 22–28. https://doi.org/10.1002/acp.3162

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

LeFevre, J., A., Greenham, S., L., & Waheed, N. (1993). The development of procedural and conceptual knowledge in computational estimation. *Cognition and Instruction, 11(2),* 95–132. https://doi.org/10.1207/s1532690xci1102_1

Levy, B. (1996). Improving memory in old age through implicit self-stereotyping. *Journal of Personality and Social Psychology, 71(6)*, 1092–1107. https://doi.org/10.1037//0022-3514.71.6.1092

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

Lemaire, P., & Brun, F. (2014a). Effects of strategy sequences and response-stimulus intervals on children's strategy selection and strategy execution: A study in computational estimation. *Psychological Research, 78(4),* 506–519. https://doi.org/10.1007/s00426-013-0501-0

Lemaire, P., & Brun, F. (2014b). Adults' age-related differences in strategy perseveration are modulated by response-stimulus intervals and problem features. *Quarterly Journal of Experimental Psychology*, *67(10),* 1863–1870. https://doi.org/10.1080/17470218.2014.939095

Lemaire, P., & Brun, F. (2016). Age–Related differences in children's strategy repetition: A study in arithmetic. *Journal of Experimental Child Psychology, 150,* 227–240. https://doi.org/10.1016/j.jecp.2016.05.014

Lemaire, P., Brun, F., Régner, I. (2018). Negative aging stereotypes disrupt both the selection and execution of strategies in older adults. *Gerontology, 64(4)*, 373-381. https://doi.org/10.1159/000486756

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., Hinault, T. (2014). Age-related differences in sequential modulations of poorer-strategy effects: a study in arithmetic problem solving. *Experimental Psychology, 61(4),* 253–262. https://doi.org/10.1027/1618-3169/a000244

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., & Lecacheur, M. (2010). Strategy switch costs in arithmetic problem solving. *Memory & Cognition, 38,* 322–332. https://doi.org/10.3758/mc.38.3.322

Lemaire, P., Leclère, M. (2014a). 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., Leclère, M. (2014b). Strategy selection in Alzheimer patients: a study in arithmetic. *Journal of Clinical and Experimental Neuropsychology, 36(5),* 507–516. https://doi.org/10.1080/13803395.2014.911248

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

Löckenhoff, C., E. (2018). Aging and Decision-Making: A Conceptual Framework for Future Research - A Mini-Review. *Gerontology, 64,* 140–148. https://doi.org/10.1159/000485247

Luchetti, M., Terracciano, A., Stephan, Y., Sutin, A. (2015). Personality and Cognitive Decline in Older Adults: Data from a Longitudinal Sample and Meta-Analysis. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 71(4),* 591–601. https://doi.org/10.1093/geronb/gbu184

Luwel, K., Onghena, P., Torbeyns, J., Schillemans, V., & Verschaffel, L. (2009). Strengths and weaknesses of the choice/no-choice method in research on strategy use. *European Psychologist, 14*(4), 351–362. http://dx.doi.org/10.1027/1016-9040.14.4.351

Masse, C., & Lemaire, P. (2001). Do people combine the parity-and five-rule checking strategies in product verification? *Psychological Research, 65(1),* 28–33. https://doi.org/10.1007/s004260000030

Margoni, F., Geipel, J., Hadjichristidis, C., Surian, L. (2018). Moral Judgement in Old Age. *Experimental Psychology, 65(2),* 105-114. https://doi.org/10.1027/1618-3169/a000395

Matthews, G. (2009). Personality and performance: Cognitive processes and models. In P.J. Corr & G. Matthews (Eds.), *The Cambridge handbook of personality psychology* (pp. 400–426). Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511596544.026

Mazerolle, M., Régner, I., Barber, S. J., Paccalin, M., Miazola, A. C., Huguet, P., & Rigalleau, F. (2017). Negative aging stereotypes impair performance on brief cognitive tests used to screen for predementia. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences,72(6)*, 932–936. https://doi.org/10.1093/geronb/gbw083

Mazerolle, M., Régner, I., Rigalleau, F., & Huguet, P. (2015). Stereotype Threat Alters the Subjective Experience of Memory. *Experimental Psychology, 62,* 395-402. https://doi.org/10.1027/1618-3169/a000303

Mazerolle, M., Régner, I., Morisset, P., Rigalleau, F., Huguet, P. (2012). Stereotype threat strengthens automatic recall and undermines controlled processes in older adults. *Psychological Science, 23,* 723–727. https://doi.org/10.1177/0956797612437607

McNair, S., Okan, Y., Hadjichristidis, C., de Bruin, W., B. (2018). Age differences in moral judgment: Older adults are more deontological than younger adults. *Journal of Behavioral Decision Making*, 1–14. https://doi.org/10.1002/bdm.2086

Moore, S., C., Oaksford, M. (2002). Some long-term effects of emotion on cognition. *British Journal of Psychology, 93(3),* 383–395. https://doi.org/10.1348/000712602760146341

Old, S., R., Naveh-Benjamin, M.(2008). Differential Effects of Age on Item and Associative Measures of Memory: A Meta-Analysis. *Psychology and Aging, 23(1),* 104–118. https://doi.org/10.1037/0882-7974.23.1.104

Osborne, J., W. (2001). Testing stereotype threat: Does anxiety explain race and sex differences in achievement? *Contemporary Educational Psychology, 26,* 291–310. https://doi.org/10.1006/ceps.2000.1052

Ownby, R., L., Crocco, E., Acevedo, A., John, V., Loewenstein, D. (2006). Depression and risk for Alzheimer disease: Systematic review, meta-analysis, and metaregression analysis. *Archives of General Psychiatry, 63(5),* 530–538. https://doi.org/10.1001/archpsyc.63.5.530

Pavlova, M., A., Wecker, M., Krombholz, K., Sokolov, A., A. (2010). Perception of intentions and actions: gender stereotype susceptibility. *Brain Research, 1311,* 81–85. https://doi.org/10.1016/j.brainres.2009.11.046

Pavlova, M., A., Weber, S., Simoes, E., & Sokolov, A., N. (2014). Gender Stereotype Susceptibility. *PLoS One, 9(12),* e114802. https://doi.org/10.1371/journal.pone.0114802

Pennington, C., R., Heim, D., Levy, A., R., & Larkin, D., T. (2016). Twenty years of stereotype threat research: A review of psychological mediators. *PLoS ONE, 11*, e0146487 https://doi.org/10.1371/journal.pone.0146487

Pessoa, L. (2012). Beyond brain regions: network perspective of cognition–emotion interactions. *Behavioral and Brain Sciences, 35(3),* 158–159. https://doi.org/10.1017/s0140525x11001567

Popham, L., E., & Hess, T., M. (2015). Age differences in the underlying mechanisms of stereotype threat effects. *Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 70B*, 223–232. https://doi.org/10.1093/geronb/gbt093

Popham, L., E., Hess, T., M. (2016) Age Stereotyping and Views of Aging, Theories of. In Pachana N. (Eds), *Encyclopedia of Geropsychology.* Singapore: Springer https://doi.org/10.1007/978-981-287-080-3_127-2

Quinn, D., M., & Spencer, S., J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues, 57(1),* 55–71. https://doi.org/10.1111/0022-4537.00201

Raven J. C. (1951). *Guide to Using Progressive Matrices* (1947), Sets A, Ab, B. London: Lewis.

Rahhal, T., A., Hasher, L., & Colcombe, S., J. (2001). Instructional manipulations and age differences in memory: Now you see them, now you don't. *Psychology and Aging, 16(4)*, 697–706. https://doi.org/10.1037//0882-7974.16.4.697

Régner, I., Smeding, A., Gimmig, D., Thinus-Blanc, C., Monteil, J., M., Huguet, P. (2010). Individual differences in working memory moderate stereotype-threat effects. *Psychological Science, 21 (11),* 1646–1648. https://doi.org/10.1177/0956797610386619

Rydell, R., J., McConnell, A., R., & Beilock, S., L. (2009). Multiple social identities and stereotype threat: Imbalance, accessibility, and working memory. *Journal of Personality and Social Psychology*, *96*, 949–966. https://doi.org/10.1037/a0014846

Royall, D., R., Palmer, R., Chiodo, L., K., Polk, M., J. (2012). Depressive symptoms predict longitudinal change in executive control but not memory. *International Journal of Geriatric Psychiatry, 27(1),* 89–96. https://doi.org/10.1002/gps.2697

Sackett, P., R., Hardison, C., M., & Cullen, M., J. (2004). On interpreting stereotype threat as accounting for African American–White differences on cognitive tests. *American Psychologist, 59,* 7–13. https://doi.org/10.1037/0003-066x.59.1.7

Salthouse, T., A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103 (3),* 403–428. https://doi.org/10.1037//0033-295x.103.3.403

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38(2)*, 194-201. https://doi.org/10.1006/jesp.2001.1500

Schmader T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85(3),* 440–452. https://doi.org/10.1037/0022-3514.85.3.440

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115(2)*, 336–356. https://doi.org/10.1037/0033-295x.115.2.336

Schmader, T., & Beilock, S. (2012). An integration of processes that underlie stereotype threat. In M. Inzlicht & T. Schmader (Eds.), *Stereotype threat: Theory, process, and application* (pp. 34-50). New York, NY, US: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199732449.003.0003

Seeman, T., E., Lusignolo, T., M., Albert, M., & Berkman, L. (2001). Social relationships, social support and patterns of cognitive aging in healthy, high-functioning older adults: MacArthur studies of successful aging. *Health Psychology, 20(4),* 243–255.

https://doi.org/10.1037//0278-6133.20.4.243

Seibt, B., & Förster, J. (2004). Stereotype threat and performance: How self-stereotypes influence processing by inducing regulatory foci. *Journal of Personality and Social Psychology, 87,* 38–56. https://doi.org/10.1037/a0013463

Siegler, R., S., & Lemaire, P. (1997). Older and younger adults' strategy choices in multiplication: Testing predictions of ASCM using the choice/no-choice method. *Journal of Experimental Psychology: General, 126(1)*, 71–92. https://doi.org/10.1037//0096-3445.126.1.71

Siegler, R., S. (2007). Cognitive variability. *Developmental Science, 10(1)*, 104–109. https://doi.org/10.1111/j.1467-7687.2007.00571.x

Smeding, A., Dumas, F., Loose, F., & Régner, I. (2013). Order of administration of math and verbal tests: An ecological intervention to reduce stereotype threat on girls' math performance. *Journal of Educational Psychology, 105(3),* 850-860. http://dx.doi.org/10.1037/a0032094

Spencer, S., J., Steele, C., M., & Quinn, D., M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35(1)*, 4–28. https://doi.org/10.1006/jesp.1998.1373

Spencer, S., J., Logel, C., & Davies, P., G. (2016). Stereotype Threat. *Annual Review of Psychology, 67(1),* 415–437. https://doi.org/10.1146/annurev-psych-073115-103235

Steele, C., M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69(5)*, 797–811. https://doi.org/10.1037//0022-3514.69.5.797

Steele, C., L., Spencer, S., J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology*, *34*, 379–440. https://doi.org/10.1016/s0065-2601(02)80009-0

Steele, C., M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52(6)*, 613-629. https://doi.org/10.1037//0003-066x.52.6.613

Stein, R., Blanchard-Fields, F. & Hertzog, C. (2002). The effects of age-stereotype priming on the memory performance of older adults. *Experimental Aging Research, 28(2),* 169-181. https://doi.org/10.1080/03610730252800184

Swift, H., J., Lamont, R., & Abrams, D. (2012). Are they half as strong as they used to be? An experiment testing whether age-related social comparisons impair older people's hand grip strength and persistence. *BMJ Open, 2(3),* 1–6. https://doi.org/10.1136/bmjopen-2012-001064

Uittenhove, K., & Lemaire, P. (2015). Numerical cognition during cognitive aging. In R. Cohen Kadosh & A. Dowker (Eds.), *Oxford Handbook of Numerical Cognition*, (pp. 345–364). Oxford: Oxford University Press. 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 aging. In Henik, A. & Fias, W. (Eds.). *Heterogeneity of function in numerical cognition*, (pp. 127–145). San Diego: Academic Press. https://doi.org/10.1016/b978-0-12-811529-9.00007-8

Voelcker-Rehage, C.(2008). Motor-skill learning in older adults—a review of studies on age-related differences. *European Review of Aging & Physical Activity, 5(1),* 5-16. https://doi.org/10.1007/s11556-008-0030-9

Wicherts, J., M. (2005). Stereotype Threat Research and the Assumptions Underlying Analysis of Covariance. *American Psychologist, 60(3),* 267-269. https://doi.org/10.1037/0003-066x.60.3.267

Wax, A. (2009). Stereotype threat: A case of overclaim syndrome? In C. H. Sommers (Ed.), *The science on women and science*, (pp. 132–169). Washington D. C.: AIE Press. https://doi.org/10.1063/1.3490501

Wechsler, D. (1997). *Administration and Scoring Manual : Wechsler Adult Intelligence Scale* (3rd ed.). San Antonio, TX: Psychological Corporation.

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). New-York and Hove: Psychology press.