When and How Stereotype Threat Influences Older Adults’ Arithmetic Performance: Insight From a Strategy Approach

Poshita Nicolas, Patrick Lemaire, and Isabelle Régner
Aix-Marseille University and CNRS, Marseille, France

In 3 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 nonexistent or very small and where effects of age-based stereotype threat have never been investigated. In 2 types of tasks (problem verification in Experiment 1 and computational estimation in Experiments 2 and 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 occur in a wide variety of cognitive domains and tasks and are independent 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

Supplemental materials: http://dx.doi.org/10.1037/xge0000647.supp

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 debated. 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
Although most ABST effects have been investigated in the domain of memory (see Armstrong, Gallant, Li, Patel, & Wong, 2017; Barber, 2017; Barber & Matner, 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 behavior (Voelcker-Rehage, 2008), skill acquisition (Head, Raz, Gunning-Dixon, Williamson, & Acker, 2002), and driving skills (Chaparro & Alton, 2005). 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 nonexistent 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 nonexistent 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, 2008; Beilock, Rydell, & McConnell, 2007; Huguet & Régner, 2007, 2009; Quinn & Spencer, 2001; Schmader, 2002; Schmader & Johns, 2003; Smeding, Dumas, Loose, & Régner, 2013; Spencer, Logel, & Davies, 2016; Spencer, Steele, & Quinn, 1999), whereas 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 were 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 Gilmore, Göbel, & Inglis, 2018; Cohen-Kadosh & Dowker, 2015, 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., Osborne, 2001; Steele, 1997; Steele, Spencer, & Aronson, 2002), depletion of executive control resources (e.g., Schmader & Johns, 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, e.g., 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. In addition, 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 (Experiment 1), the other two in which they were investigated directly (Experiments 2 and 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; Experiment 1), strategies that people use cannot be directly observed (i.e., with external behavioral 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; Experiments 2 and 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 non-threat and threat conditions (Bruhuber & 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), 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., 2008; Abrams et al., 2006). 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 more 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 in 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 × 17 = 68, True or False?). Previous studies found no or small age-related differences on participants’ performance in arithmetic problem verification tasks (e.g., Allen, Ashcraft, & Weber, 1992; Allen, Smith, Jerge, & Vires-Collins, 1997; Duverne & Lemaire, 2004). Participants were asked to verify true (e.g., 4 × 17 = 68) and false (e.g., 4 × 13 = 41) equations. True problems were either small problems (e.g., 4 × 18 = 72) or large problems (e.g., 8 × 49 = 392). We tested whether ABST would change so-called problem size effects (i.e., poorer performance on large problems than on small problems, one of the most robust arithmetic effects; see Zbrodoff & Logan, 2005, for a review).

Also, participants verified false problems for which the parity of the proposed answer either matched (e.g., 19 × 7 = 131) or mismatched (e.g., 19 × 7 = 134) that of the correct answer. This manipulation was based on previous studies showing that young and older participants obtain better performance when the parity of proposed and correct answer differs relative to when the parity is the same (Anders, Hinault, & Lemaire, 2018; Hinault & Lemaire, 2017; Hinault, Dufau, & Lemaire, 2015; Hinault, Tiberghien, & Lemaire, 2016; Krueger & Hallford, 1984; Krueger, 1986; Lemaire & Reder, 1999; Lemaire & Fayol, 1995; Masse & Lemaire, 2001). The goal was to determine whether ABST changes magnitudes of these so-called parity violation effects. Parity violation
effects have been accounted for by assuming that participants use
two different strategies, a slower exhaustive-verification strategy
to verify false, parity-match problems, and a faster plausibility-
checking strategy on parity-mismatch problems. In the exhaustive-
verification strategy, participants encode the problem, search for
the correct answer, compare the proposed and correct answers,
make a true/false decision, and provide their response. In the plausibility-checking strategy, they quickly decide to reject the
equation without searching for the correct answer given the dif-
ference in parity between the proposed and correct answers.
Changes in magnitudes of parity violation effects under stereotype
threat would suggest that stereotype threat influences strategies
that participants use on each problem.

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 accom-
plished the Mini Mental-State Examination (MMSE; Folstein,
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) × 2 (group: threat, control) 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-violation effects, which both involve probing for a between-within interaction (that typically requires a smaller sample size). To achieve larger than 80% power, we decided to recruit 40 participants per group, resulting in a final sample of 160 participants.

Stimuli. The stimuli were 288 multiplication problems pre-
sent in a standard form (i.e., $a \times b = c$), with the operands $a$ and $b$ being either single- or double-digit numbers. All problems included one two-digit operand and one one-digit operand (e.g., $19 \times 7 = 131$). Single-digit operands ranged from 3 to 9, whereas
double-digit operands ranged from 12 to 97. Two types of multi-
plication 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 (e.g., $8 \times 9 = 392$) and half were small problems (e.g., $4 \times 18 = 72$). Correct products ranged from 48 to 204 for small problems ($M = 139$) and from 207 to 441 for large problems ($M = 301$).

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., $19 \times 7 = 131$), and (b) parity-mismatch (or mismatch) problems involved false answers with odd-even status that were different from those of correct products (e.g., $19 \times 7 = 132$). All false problems were created by varying splits (i.e., differences) between correct and proposed products. Incorrect answers were off by $\pm 1$, $\pm 2$, or $\pm 3$ 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, each six each from splits of $\pm 1$ and $\pm 3$ for mismatch problems, and 12 each from splits of $\pm 2$ for
match problems.

Based on previous findings in arithmetic (see Gilmore et al.,
2018; Cohen-Kadosh & Dowker, 2015, for overviews), we con-
trolled 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., $17 \times 6 = 102$), and half in the right
position (e.g., $4 \times 17 = 68$); (c) all problems with only one even
operand had half of their even multiplicant in the right position
(e.g., $17 \times 4 = 68$) and the other half in the left position (e.g., $6 \times 13 = 78$); (d) none of the problems included zero, one, or five as a
single-digit operand; and (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).

Procedure. Participants were tested individually in one ses-
sion that lasted approximately 90 min. Half the participants were

<table>
<thead>
<tr>
<th>Variables</th>
<th>Young adults</th>
<th></th>
<th>Olders adults</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Threat</td>
<td>Control</td>
<td>Threat</td>
<td>Control</td>
</tr>
<tr>
<td></td>
<td>$M (SD)$</td>
<td>$M (SD)$</td>
<td>$M (SD)$</td>
<td>$M (SD)$</td>
</tr>
<tr>
<td>n (females)</td>
<td>40 (30)</td>
<td>40 (28)</td>
<td>40 (28)</td>
<td>40 (26)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>21 (3.4)</td>
<td>20.2 (2)</td>
<td>70 (5.3)</td>
<td>69 (3.5)</td>
</tr>
<tr>
<td>Education (years)</td>
<td>14 (1.7)</td>
<td>13.5 (1.1)</td>
<td>14.6 (1.5)</td>
<td>14.3 (1.9)</td>
</tr>
<tr>
<td>Mini Mental-State Examination</td>
<td>30.5 (10.5)</td>
<td>25.7 (12.7)</td>
<td>29.5 (10.5)</td>
<td>29.2 (8)</td>
</tr>
<tr>
<td>Arithmetic Fluency Test</td>
<td>22 (4.6)</td>
<td>21.3 (3.7)</td>
<td>72.5 (11.9)</td>
<td>86.9 (26.8)</td>
</tr>
<tr>
<td>Mill-Hill Vocabulary Scale</td>
<td></td>
<td></td>
<td>27.8 (4.5)</td>
<td>27.7 (3.4)</td>
</tr>
</tbody>
</table>

Table 1
Participants’ Characteristics in Threat and Control Groups
randomly assigned to the Threat group and half to the Control group.

**Threat group.** Our stereotype threat manipulation was adapted from previous studies (e.g., Mazeroille 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., $17 \times 6 = 102$) 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 min 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., $17 \times 6 = 102$) 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 $1366 \times 768$ 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 center of the screen in a standard arithmetic format $a \times b = c$. 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 center 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 from the analyses because they made 52% and 53% errors, respectively.

**Effects of ABST on false problems performance.** Mean response latencies and percentages of errors on false problems were analyzed with 2 (age: young, older adults) $\times 2$ (group: threat, control) $\times 2$ (parity: match, mismatch problems) mixed-design analyses of variance (ANOVs), with age and group as the between-participants factors (see means in Table 2 and summary of significant effects on latencies in Table 3).

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 $\times$ Group interaction showed that young adults tested in the threat group and in the control group were equally fast ($F < 1$), whereas 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 $\times$ Group $\times$ 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

---

1 Analyses including and excluding these participants yielded the same results in 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

<table>
<thead>
<tr>
<th>Problems</th>
<th>Threat</th>
<th>Control</th>
<th>Ms</th>
<th>Threat</th>
<th>Control</th>
<th>Ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Match</td>
<td>3,940 (11.4)</td>
<td>3,871 (10.3)</td>
<td>3,905 (10.8)</td>
<td>4,239 (6.0)</td>
<td>3,891 (6.0)</td>
<td>4,065 (6.0)</td>
</tr>
<tr>
<td>Mismatch</td>
<td>3,511 (7.0)</td>
<td>3,345 (5.7)</td>
<td>3,428 (6.3)</td>
<td>4,159 (3.0)</td>
<td>3,461 (2.3)</td>
<td>3,810 (2.7)</td>
</tr>
<tr>
<td>PVE</td>
<td>429* (4.9)</td>
<td>526** (4.6)</td>
<td>477 (4.5)</td>
<td>80 (3.3)</td>
<td>430** (3.7)</td>
<td>255 (3.3)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Problems</th>
<th>Threat</th>
<th>Control</th>
<th>Ms</th>
<th>Threat</th>
<th>Control</th>
<th>Ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large</td>
<td>4,548 (17.7)</td>
<td>4,461 (15.9)</td>
<td>4,504 (16.8)</td>
<td>5,314 (11.3)</td>
<td>4,550 (9.9)</td>
<td>4,922 (10.6)</td>
</tr>
<tr>
<td>Small</td>
<td>4,262 (11.7)</td>
<td>4,218 (10.1)</td>
<td>4,240 (10.9)</td>
<td>4,673 (8.2)</td>
<td>4,197 (8.7)</td>
<td>4,438 (8.5)</td>
</tr>
<tr>
<td>M</td>
<td>4,405 (14.7)</td>
<td>4,339 (13)</td>
<td>4,372 (13.8)</td>
<td>4,993 (9.7)</td>
<td>4,373 (9.3)</td>
<td>4,683 (9.6)</td>
</tr>
<tr>
<td>PSE</td>
<td>286* (6.0)</td>
<td>243** (5.8)</td>
<td>264 (5.9)</td>
<td>641* (3.1)</td>
<td>353** (1.2)</td>
<td>497 (2.1)</td>
</tr>
</tbody>
</table>

Note.  PVE (parity violation effects) = match − mismatch; PSE (problem size effects) = large − small.  **p < .01.

(Threat−Control = 348 ms; F[1, 154] = 8.56, MS e = 2,420,340, ƞ2 = 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 large problems (299 ms; F[1, 154] = 8.56, MS e = 1,739,368, ƞ2 = 0.04) and mismatch problems (648 ms; F[1, 154] = 26.25, MS e = 8,188,926, ƞ2 = 0.15). In contrast, young and older adults in the control group had comparable performance on both match and mismatch problems (Fs < 1). Parity violation effects (see Figure 1) were significant in young adults, both in the threat group (match−mismatch = 429 ms; F[1, 154] = 54.70, ƞ2 = 0.26) and in the control group (match−mismatch = 526 ms; F[1, 154] = 86.40, ƞ2 = 0.36). However, no parity violation effects were found in older participants tested in the threat group (match−mismatch = 80 ms; F < 2) in contrast to significant parity violation effects in older participants tested under the control group (match−mismatch = 430 ms; F[1, 154] = 57.71, ƞ2 = 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, MS e = 721, ƞ2 = 0.14. Participants made more errors on match than on mismatch problems (8.4% vs. 4.5%), F(1, 154) = 94.03, MS e = 1,217, ƞ2 = 0.38. No other effects came out significant on either latencies or errors.

Effects of ABST on true problems performance. Participants’ performance on true problems were analyzed with 2 (age: young, older adults) × 2 (group: threat, control) × 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 × 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, MS e = 7,670,959, ƞ2 = 0.12. Most interestingly, the Age × Group × 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, MS e = 11,650,248, ƞ2 = 0.14) than while verifying true small problems (threat−control = 476 ms; F[1, 154] = 12.85, MS e = 4,520,103, ƞ2 = 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 (764 ms; F[1, 154] = 476 ms; F[1, 154] = 9.35, MS e = 3,289,539, ƞ2 = 0.06). In contrast, young and

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

<table>
<thead>
<tr>
<th>Effects</th>
<th>df</th>
<th>MS e</th>
<th>F</th>
<th>ƞ2</th>
</tr>
</thead>
<tbody>
<tr>
<td>False problems</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>1,154</td>
<td>4,039,622</td>
<td>15.22***</td>
<td>0.09</td>
</tr>
<tr>
<td>Age</td>
<td>1,154</td>
<td>2,888,820</td>
<td>10.88***</td>
<td>0.07</td>
</tr>
<tr>
<td>Parity</td>
<td>1,154</td>
<td>10,583,599</td>
<td>165.39***</td>
<td>0.52</td>
</tr>
<tr>
<td>Age × Group</td>
<td>1,154</td>
<td>1,026,968</td>
<td>6.13*</td>
<td>0.04</td>
</tr>
<tr>
<td>Age × Parity</td>
<td>1,154</td>
<td>979,619</td>
<td>15.31***</td>
<td>0.09</td>
</tr>
<tr>
<td>Group × Parity</td>
<td>1,154</td>
<td>984,149</td>
<td>15.38***</td>
<td>0.09</td>
</tr>
<tr>
<td>Age × Group × Parity</td>
<td>1,154</td>
<td>316,977</td>
<td>4.95*</td>
<td>0.03</td>
</tr>
</tbody>
</table>

| 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 × Group      | 1,154    | 3,029,754  | 8.18**  | 0.05 |
| Age × Problem Size | 1,154 | 1,068,068 | 12.81*** | 0.08 |
| Group × Problem Size | 1,154 | 538,521   | 6.46**  | 0.04 |
| Age × Group × Problem Size | 1,154 | 296,988   | 3.56**  | 0.02 |

*p < .05. **p < .01. ***p < .001.
older adults in the control group had comparable performance on both large and small problems ($F s < 1$). Problem-size effects (see Figure 1) were of the same magnitudes in young adults tested in the threat group (large–small = 286 ms; $F[1, 154] = 18.60, \eta^2 = 0.11$) and in the control group (large–small = 243 ms; $F[1, 154] = 14.19, \eta^2 = 0.08$). However, larger problem-size effects were found for older adults in the threat group (large–small = 641 ms; $F[1, 154] = 98.55, \eta^2 = 0.39$) compared to the control group (large–small = 353 ms; $F[1, 154] = 29.92, \eta^2 = 0.16$).

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

**Effects of ABST 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) × 2 (group: threat, control) between-participants ANOVA. Results showed a main effect of age on arithmetic fluency scores, $F(1, 154) = 367.4, MSe = 105,073, \eta^2 = 0.71$ (i.e., older adults scored higher than young adults on arithmetic fluency test, 79.6 vs. 28.1). Interestingly, the Age × Group interaction was significant, $F(1, 154) = 12.62, MSe = 3,609, \eta^2 = 0.08$. 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; $F[1, 154] = 14.55, MSe = 4,161, \eta^2 = 0.09$) and of equal performance in both groups of young adults ($F s < 2$).

**Experiment 2**

In Experiment 1, we found adverse effects of ABST on older adults’ arithmetic performance. These effects were found in arithmetic verification task where age-related differences are inexistent. We found decreased performance in older adults only under stereotype threat condition relative to control condition where both young and older adults showed comparable performance. More specifically, we found increased problem size effects and decreased parity violation effects. Following previous works in arithmetic (see Cohen-Kadosh & Dowker, 2015; Gilmore et al., 2018, for overviews), increased problem size effects could be interpreted as resulting from older adults executing the exhaustive-verification strategy more slowly on large problems under stereotype threat. In other words, stereotype threat may have 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 occurred because participants did not use as often and as systematically the fast plausibility-checking strategy (based on the discrepancy between the parity of proposed and correct answers) on false mismatch problems but used the slower exhaustive-verification strategy under stereotype threat.

Although results from Experiment 1 indicates that ABST effects are independent of older adults’ baseline performance, a limitation of this experiment concerns lack of direct evidence for the strategy hypothesis that ABST effects occur via strategic variations. Patterns of participants’ performance as a function of stimulus features (problem size, parity) suggest that ABST influences which strategy participants used and how they executed the selected strategy on each problem. But verification tasks do not offer the possibility to investigate strategies directly. The second limitation of Experiment 1 for the strategy hypothesis is that 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 $32 \times 79$). Numerous previous studies used this task to investigate strategic aspects of arithmetic problem solving and age-related differences therein (e.g., Hinault & Lemaire, 2017; LeFevre, Greenham, & Waheed, 1993; Lemaire & Leclère, 2014a). We also used the choice/no-choice method proposed by Siegler and Lemaire (1997) that has been 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 estimates 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 in 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., $63 \times 82$). 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, Dufau, & Lemaire, 2014; Hinault & Lemaire, 2017; Lemaire & Hinault, 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 MMSE (Folstein et al., 1975), and none obtained a MMSE score <.27. Participants also completed a French version of the 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) × 2 (trials: repeated, unrepeated) interaction to assess the influence of ABST on strategy repetitions in the choice condition, and the 2 (group: threat, control) × 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 $d = .52$ (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.

Stimuli. The stimuli for the computational estimation task were 64 two-digit multiplication problems presented in a standard form (i.e., $a \times b$; $32 \times 67$). Based on the size of the unit digits, half the problems were so-called homogeneous problems, and half were so-called heterogeneous problems. Unit digits of both operands were either smaller than five (e.g., $32 \times 63$) or larger than five (e.g., $38 \times 69$) 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., $49 \times 62$) for half the problems and the reverse (e.g., $42 \times 58$) 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 $83 \times 21$, and a heterogeneous problem, like $86 \times 21$, are best estimated with the rounding-down strategy, doing $80 \times 20 = 1,600$). The other problems were best estimated with the rounding-up strategy (e.g., a homogeneous problem, like $78 \times 39$, and a heterogeneous problem, like $72 \times 39$, are best estimated with the rounding-up strategy, doing $80 \times 40 = 3,200$).

Table 4

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Threat $M$ (SD)</th>
<th>Control $M$ (SD)</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$ (females)</td>
<td>28 (22)</td>
<td>28 (19)</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>73.2 (5.4)</td>
<td>72.6 (5.5)</td>
<td>.69</td>
</tr>
<tr>
<td>Education (years)</td>
<td>13.1 (2.5)</td>
<td>12 (1.6)</td>
<td>.07</td>
</tr>
<tr>
<td>MMSE</td>
<td>29 (.8)</td>
<td>28.9 (.8)</td>
<td>.09</td>
</tr>
<tr>
<td>MHVS</td>
<td>28 (3.1)</td>
<td>26 (5.4)</td>
<td>.07</td>
</tr>
</tbody>
</table>

Note. MMSE = Mini Mental-State Examination; MHVS = Mill-Hill Vocabulary Scale.
We tested homogeneous and heterogeneous problems because previous studies found that participants select the best rounding strategy more easily on homogeneous than on heterogeneous problems (e.g., LeFevre et al., 1993; Lemaire, Arnaud, & Lecacheur, 2004). Homogeneous and heterogeneous problems had comparable exact products when solved with each rounding strategy (means were 2,426 and 2,408 for homogeneous and heterogeneous problems, respectively).

Finally, following previous studies on strategy repetition (e.g., Lemaire & Brun, 2016; Lemaire & Leclère, 2014a, 2014b), 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 un-repeated-strategy trials (two successive problems were best estimated using different strategies).

Given well-known effects in the domain of mental arithmetic (see Gilmore et al., 2018; Cohen-Kadosh & Dowker, 2015, 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., $33 \times 42$), (c) digits were not repeated in the same unit or decade positions across operands (e.g., $62 \times 67$), (d) no reverse order of operands was used (e.g., $36 \times 23$ vs. $23 \times 36$), (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., $42 \times 36$) and in the right position in the other problems (e.g., $23 \times 41$).

**Procedure.** Participants were tested in one session that lasted approximately 60 min. 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 1,366 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 center of the screen in a standard arithmetic format « $a \times b$ ». 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; $80 \times 20$ to estimate $86 \times 21$) or the rounding-up strategy (i.e., rounding the two operands up to the nearest decades; $90 \times 30$ to estimate $87 \times 26$). They were also told that no other strategy could be used (e.g., mixed rounding strategy, like rounding the first operand down and the second operand up to the closest decades, such as when doing $80 \times 30$ to estimate $84 \times 29$). 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). Previous studies showed that both young and older adults are able to use these mixed-rounding strategies, and that when mixed-rounding strategies are available together with rounding-up/down strategies, both young and older adults choose the better strategy on over 95% of problems (e.g., LeFevre et al., 1993; Lemaire et al., 2004). Inclusion of mixed-rounding strategies may have therefore resulted in all participants (in both threat and control groups) using the best strategy on almost all the problems. This would make the task less sensitive to any threat effects on strategy use. 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 $56 \times 78$), and the poorer strategy was required on the other problems (e.g., rounding down was cued for $57 \times 68$). 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, & Phillips, 2016; Hinault, Lemaire, & Touron, 2017; 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 min 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 ABST first on strategy selection and participants’ performance...
in the Choice condition, and second on strategy execution in the No-Choice condition.

**Effects of ABST on strategy selection and performance in the choice condition.**

**Strategy selection.** We first conducted a 2 (group: threat, control) × 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).

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, ηp² = 0.68) and in the control than in the threat group, F(1, 54) = 49.2, MSe = 6,246, ηp² = 0.48). The significant Group × Problem interaction, F(1, 54) = 10.68, MSe = 846, ηp² = 0.16, showed that the effect of threat was larger on homogeneous problems (control–threat = 27%; F[1, 54] = 51.8, MSe = 9,922, ηp² = 0.49) than on heterogeneous problems (control–threat = 15%; F[1, 54] = 24.1, MSe = 3,418, ηp² = 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) × 2 trial (repeated, unrepeated) design ANOVA on mean percentages of use of the better strategy, 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, ηp² = 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, ηp² = 0.30. Most importantly, the Group × Trial interaction (see Figure 2), F(1, 54) = 55.4, MSe = 12,118, ηp² = 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, ηp² = 0.08), and the reverse for unrepeated trials, (30% vs. 66%; F[1, 54] = 46.8, MSe = 17,839, ηp² = 0.46).

**Participants’ performance.** We carried out 2 (group: threat, control) × 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, ηp² = 0.42. Also, participants in the threat group were faster than participants in the control group, F(1, 54) = 7.7, MSe = 15,090,963, ηp² = 0.13. The Group × 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, ηp² = 0.21. In other words, participants in the threat group were 1,438 ms faster than participants in the control group on homogeneous problems but equally fast on heterogeneous problems (F < 3.0). Interestingly, analyses of 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, ηp² = 0.23) and more on heterogeneous than on homogeneous problems (mean difference = 1.9; F[1, 54] = 6.30, MSe = 106, ηp² = 0.10). The Group × Problem interaction revealed significant differences between the threat and the control groups on homogeneous problems only, F(1, 54) = 43.65, MSe = 1,253, ηp² = 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 (CI) 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 = .6). Percentages of use of the better

**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**

<table>
<thead>
<tr>
<th>Problems</th>
<th>Threat</th>
<th>Control</th>
<th>Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homogeneous</td>
<td>67</td>
<td>94</td>
<td>80</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>55</td>
<td>70</td>
<td>62</td>
</tr>
<tr>
<td>M</td>
<td>61</td>
<td>82</td>
<td>71</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Problems</th>
<th>Mean solution latencies in ms (and percent deviations)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homogeneous</td>
<td>4,088 (11.3) 5,526 (1.9) 4,807 (6.6)</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>5,091 (9.4) 5,730 (7.7) 5,410 (8.6)</td>
</tr>
<tr>
<td>M</td>
<td>4,590 (10.3) 5,628 (4.8) 5,109 (7.6)</td>
</tr>
</tbody>
</table>

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) × 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, ηp² = 0.27).
strategy was thus a significant mediator of stereotype threat on percent deviations.

**Effects of ABST on strategy performance in the no-choice condition.** We conducted $2$ (group: threat, control) × $2$ (problem: homogeneous, heterogeneous problems) × $2$ (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).

Participants were slower while executing the poorer-strategy relative to while executing the better strategy on a given problem ($4,791$ ms vs. $5,590$ ms; $F(1, 54) = 161.07, MSe = 35,745,662, \eta_p^2 = 0.75$). The significant Group × Trial interaction revealed significant differences between participants’ latencies in the threat and control groups on poorer-strategy trials only, $F(1, 54) = 8.8, MSe = 14,781,615, \eta_p^2 = 0.14$. In other words, older adults were slower on poorer strategy trials when tested under stereotype threat than when tested under control group ($M$ difference = $1,028$ ms). Older adults had comparable performance under threat and control groups on better-strategy trials ($M$ difference = $40$ ms; $F < 1.0$). The Group × Trial × Problem interaction was not significant, $F < 1.0$.

Analyses of mean absolute percent deviations revealed a significant Group × Trial interaction. This resulted from differences between the threat and the control groups being significant on better-strategy trials only, $F(1, 54) = 4.8, MSe = 22, \eta_p^2 = 0.08$. 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 (1997), threat can be elicited within the same individual just by changing the situational aspect.

---

**Figure 2.** Older adults’ mean percentages of strategy repetition on repeated trials (when two successive problems were of the same type) and unrepeated trials (when two successive problems were of different types) in the threat and control groups. Error Bars represent $95\%$ confidence intervals.

**Figure 3.** Simple mediation model for the percentages of use of the better strategy on absolute percent deviations.
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 pretest and posttest 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 MHVS (Deltour, 1993; Raven, 1951). None obtained a 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 2 × 2 mixed design to achieve at least 80% power to detect the critical 2(Group: threat, control) × 2(Session: pretest, posttest) interaction. To calculate this sample size, we used the same parameters as used in Experiment 2 (i.e., d = 0.52, and r = .30) for the same reasons.

Processing speed measures. The WAIS-IV Digit symbol coding task (Codes) and the Choice reaction time (RT) 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 RT 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 (i.e., a × b; 34 × 67). 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 × 27 and 48 × 79, followed by a target problem, 46 × 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 × 64 and 37 × 69). These problems were chosen as priming problems because previous works (e.g., LeFevre et al., 1993; Lemaire & Brun, 2014a, 2014b) showed that both young and older adults

Table 6

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Mean (SD)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>72 (4.6)</td>
<td>.12</td>
</tr>
<tr>
<td>Education (years)</td>
<td>13 (1.9)</td>
<td>.38</td>
</tr>
<tr>
<td>MMSE</td>
<td>29 (.8)</td>
<td>.27</td>
</tr>
<tr>
<td>MHVS</td>
<td>22.3 (5.6)</td>
<td>.12</td>
</tr>
</tbody>
</table>

Note. MMSE = Mini Mental-State Examination; MHVS = Mill-Hill Vocabulary Scale.
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 × 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, 2014b) that revealed the appropriateness of this 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, 2014b; Lemaire & Leclère, 2014a, 2014b), 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 Cohen-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 × 6, 25 × 63); (b) no digits were repeated within operands (e.g., 22 × 63); (c) the first operand was larger than the second operand in half the problems, and vice versa; (d) no operand had its closest decade equal to 0, 10, or 100; and (e) rounded operands were never the same across two successive problems in a given trial (e.g., if one problem in a trial was 32 × 64, the next problem could not be 31 × 62).

Procedure. The participants were individually tested in two sessions, each of which lasted approximately 45–60 min. 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: MHVS, Codes, Color-Detection, and the French Kit. In Session 2, all participants completed the computational estimation task twice, as pretest and posttest. In both pre- and posttests, the two-digit multiplication problems were presented horizontally in a 84 point Bold Courier font (black color) on a 1,366 × 768 resolution screen. The E-Prime software 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 posttest 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 × 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 × 40 to solve 78 × 43) and the rounding-up strategy as rounding both operands up to the nearest decades (e.g., 80 × 50 to solve 78 × 43). Further, the computational estimation task was presented as comprising two blocks of 24 trials each. The two blocks were actually the pretest and the posttest administered one after the other with a brief rest period of 3–5 min between the two tests. The pretest and the posttest 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 × 27 in the pretest was 27 × 78 in the posttest), (b) of the 24 trials, the last 12 trials of the pretest were presented first in the posttest, and vice versa, and (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 pretest 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 × 26, using either a rounding-up or a rounding-down strategy). This followed a brief rest period of 3–5 min 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 posttest 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 posttest.

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 center 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., tric) or only vowels (e.g., aeo), and on the “S” key when the four letters included both consonants and vowels (e.g., ubqi). A blank screen was finally displayed for 1,000 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 ABST 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) × 2 (session: pretest, posttest) mixed-design ANOVAs, with repeated measures on the last factor.

Strategy performance on priming problems. Participants were overall faster in the posttest compared to the pretest session (pretest–posttest = 1249 ms; $F(1, 46) = 83.75, \, MS_e = 37,845,001, \, \eta^2_p = 0.57$). The Group × Session interaction was significant, $F(1, 46) = 7.37, \, MS_e = 3,300,342, \, \eta^2_p = 0.14$. This interaction resulted from larger differences in estimation latencies between the pretest and posttest sessions in the threat group (pretest–posttest = 1,620 ms; $F(1, 46) = 70.41, \, \eta^2_p = 0.60$) than in the control group (pretest–posttest = 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%; $F(1, 46) = 78.93, \, MS_e = 135, \, \eta^2_p = 0.63$), and in the posttest than in the pretest session (posttest–pretest = 2.6%; $F(1, 46) = 45.52, \, MS_e = 156, \, \eta^2_p = 0.49$). The Group × Session interaction came out significant, $F(1, 46) = 55.68, \, MS_e = 191, \, \eta^2_p = 0.55$. Further contrasts revealed that the difference in percent deviations between the threat and control groups was significant in the posttest session (threat–control = 6.2%; $F(1, 46) = 79.19, \, MS_e = 459, \, \eta^2_p = 0.63$), but not in the pretest session (threat–control = 0.5%; $F < 3$). Comparing the pretest and posttest sessions within each group showed that the difference in percent deviations between the two sessions was significant in the threat group (posttest–pretest = 5.4%; $F(1, 23) = 53.22, \, MS_e = 347, \, \eta^2_p = 0.69$), but not in the control group (pretest–posttest = 0.3%; $F < 2$).

Strategy performance on target problems. Participants were overall faster in the posttest compared to the pretest session (pretest–posttest = 1080 ms; $F(1, 46) = 40.47, \, MS_e = 27,951,819, \, \eta^2_p = 0.47$). The Group × Session interaction, $F(1, 46) = 5.17, \, MS_e = 3,576,328, \, \eta^2_p = 0.1$, was significant. This interaction resulted from larger differences in estimation latencies between the pretest and posttest sessions in the threat group (pretest–posttest = 1,465 ms; $F(1, 46) = 37.29, \, \eta^2_p = 0.45$) than in the control group (pretest–posttest = 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%; $F(1, 46) = 18.09, \, MS_e = 50, \, \eta^2_p = 0.28$), and in the posttest than in the pretest session (posttest–pretest = 1.0%; $F(1, 46) = 8.90, \, MS_e = 24, \, \eta^2_p = 0.16$). The Group × Session interaction was significant, $F(1, 46) = 7.21, \, MS_e = 19, \, \eta^2_p = 0.14$. Difference in percent deviations between the threat and control groups was significant in the posttest session (threat–control = 3.0%; $F(1, 46) = 33.90, \, MS_e = 104, \, \eta^2_p = 0.42$), but not in the pretest session (threat–control = 1.2%; $F < 3$). Comparing the pretest and posttest sessions within each group showed that difference in percent deviations between the two sessions was significant in the threat group (posttest–pretest = 1.9%; $F(1, 23) = 8.36, \, MS_e = 87, \, \eta^2_p = 0.27$), but not in the control group (pretest–posttest = 0.1%; $F < 1$).

Effects of ABST on participants’ strategy selection. Mean percentages of use of the better strategy (see means in Table 8) were

Table 8

<table>
<thead>
<tr>
<th>Sessions</th>
<th>Threat</th>
<th>Control</th>
<th>$M$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Priming problems</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>96</td>
<td>95</td>
<td>95</td>
</tr>
<tr>
<td>Posttest</td>
<td>68</td>
<td>97</td>
<td>82</td>
</tr>
<tr>
<td>$M$</td>
<td>82</td>
<td>96</td>
<td>89</td>
</tr>
<tr>
<td>Mean estimation latencies in ms (and percent deviations)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>5,988 (10.4)</td>
<td>5,495 (9.9)</td>
<td>5,741 (10.1)</td>
</tr>
<tr>
<td>Posttest</td>
<td>4,368 (15.8)</td>
<td>4,616 (9.6)</td>
<td>4,492 (12.7)</td>
</tr>
<tr>
<td>$M$</td>
<td>5,178 (13.1)</td>
<td>5,055 (9.7)</td>
<td>5,116 (11.4)</td>
</tr>
<tr>
<td><strong>Target problems</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>84</td>
<td>83</td>
<td>83</td>
</tr>
<tr>
<td>Posttest</td>
<td>56</td>
<td>84</td>
<td>70</td>
</tr>
<tr>
<td>$M$</td>
<td>70</td>
<td>83</td>
<td>76</td>
</tr>
<tr>
<td>Mean estimation latencies in ms (and percent deviations)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>6,883 (16.5)</td>
<td>6,133 (15.3)</td>
<td>6,508 (15.9)</td>
</tr>
<tr>
<td>Posttest</td>
<td>5,417 (18.4)</td>
<td>5,440 (15.4)</td>
<td>5,428 (16.9)</td>
</tr>
<tr>
<td>$M$</td>
<td>6,150 (17.5)</td>
<td>5,786 (15.4)</td>
<td>5,968 (16.4)</td>
</tr>
</tbody>
</table>
analyzed by 2 (group: threat, control) × 2 (session: pretest, posttest) 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%; \(F[1, 46] = 57.03, MSe = 2359, \eta^2_p = 0.55\)), and in the pretest compared to the posttest session (pretest–posttest = 13%; \(F[1, 46] = 43.82, MSe = 4017, \eta^2_p = 0.48\)). The Group × Session interaction was significant, \(F(1, 46) = 53.64, MSe = 4916, \eta^2_p = 0.54\). The difference in percentages of use of the better strategy between the control and the threat groups was significant in the posttest session (control–threat = 29%; \(F[1, 46] = 61.33, MSe = 9633, \eta^2_p = 0.57\), but not in the pretest session (threat–control = 1%; \(F < 1\)). Comparing the pretest and the posttest 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 (pretest–posttest = 28%; \(F[1, 23] = 51.96, MSe = 8910, \eta^2_p = 0.69\) but not in the control group (posttest–pretest = −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%; \(F[1, 46] = 51.07, MSe = 2193, \eta^2_p = 0.53\)), and in the pretest compared to the posttest session (pretest–posttest = 13%; \(F[1, 46] = 55.13, MSe = 4280, \eta^2_p = 0.54\)). The Group × Session interaction was significant, \(F(1, 46) = 65.18, MSe = 5060, \eta^2_p = 0.59\). The difference in percentages of use of the better strategy between the control and the threat groups was significant in the posttest session (control–threat = 28%; \(F[1, 46] = 94.47, MSe = 9436, \eta^2_p = 0.67\), but not in the pretest session (threat–control = 1%; \(F < 1\)). Comparing the pretest and the posttest 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 (pretest–posttest = 28%; \(F[1, 23] = 121.41, MSe = 9324, \eta^2_p = 0.84\) but not in the control group (posttest–pretest = 1%; \(F < 1\)).

### Strategy repetitions. We conducted a 2 (group: threat, control) × 2 (session: pretest, posttest) × 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.

Participants repeated the same strategy across two successive problems more often on repeated trials than on unrepeated trials (repeated−unrepeated = 53%; \(F[1, 46] = 810.86, MSe = 137.067, \eta^2_p = 0.94\)). Most interestingly, the Group × Session × Trial interaction was significant, \(F(1, 46) = 65.37, MSe = 10150, \eta^2_p = 0.58\). During posttest, 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%; \(F[1, 46] = 31.31, MSe = 6745, \eta^2_p = 0.41\) and less often in the control than in the threat condition for unrepeated trials (control–threat = −32.3%; \(F[1, 46] = 43.95, MSe = 12545, \eta^2_p = 0.49\)). No differences were found between threat and control groups in the pretest session for either repeated or unrepeated trials (\(Fs < 1.0\)). Comparing the pretest and the posttest 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 (pretest–posttest = 22.3%; \(F[1, 23] = 17, \eta^2_p = 0.42\)) and unrepeated trials (pretest–posttest = 33.2%; \(F[1, 23] = 70, \eta^2_p = 0.75\)). No significant differences in percentages of strategy repetitions were found between the two sessions in the control group (\(Fs < 2\)).

**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 \(ps > .7\).

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 (5,000 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).

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 × Arithmetic Fluency and Group × 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 × Processing Speed\(^3\) was not significant while controlling for Threat × Arithmetic Fluency; \(F(1, 42) = 0.38; \Delta R^2 = .002; p = .536\)). Thus, we found that arithmetic fluency was the best moderator. To interpret this moderation effect, we further exam-

<table>
<thead>
<tr>
<th>Session</th>
<th>Threat</th>
<th>Control</th>
<th>(M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>90</td>
<td>87</td>
<td>88</td>
</tr>
<tr>
<td>Posttest</td>
<td>68</td>
<td>91</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>79</td>
<td>89</td>
<td>84</td>
</tr>
<tr>
<td>Unrepeated trials</td>
<td>22</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>Posttest</td>
<td>56</td>
<td>23</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>23</td>
<td>31</td>
</tr>
</tbody>
</table>

\(^3\) 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 (\(r = −.371, p = .004\)). We therefore conducted a simple moderation analysis (Model 1) and found a significant Group × Processing Speed interaction suggesting that the relation between threat and percent deviations was moderated by processing speed, \(F(1, 44) = 5.78; \Delta R^2 = .049; p = .020\). It would be interesting for future studies to investigate processing speed as a moderator of threat effects in domains other than arithmetic.
ined 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 ($b_{11005}/H11005/H11002 = 2.93, 95\% \text{ CI } [3.64, 2.21], SE_{11005} = 0.355, t_{11005} = 8.25, p_{11021} = .001$) to a smaller negative effect ($b_{11005/H11002} = 0.718, 95\% \text{ CI } [1.36, 0.07], SE_{11005} = 0.320, t_{11005} = 2.24, p_{11005} = .030$).

**General Discussion**

The three experiments reported here reflected the convergence of several unique 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 (Experiments 1 and 2) and within-participants (Experiment 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, pretest–posttest 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. The findings obtained from the analytical approach used across experiments (to best analyze our data as function of the hypotheses under test and for the purposes of comparison with previous data in the arithmetic literature) are not inconsistent with other analytical approaches (see online supplemental materials). Our results establish that ABST effects can impair older adults’ performance even on tasks where they are competent (i.e., ABST effects are independent of any preexperimental 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.

---

**Table 10**

<table>
<thead>
<tr>
<th>Linear Models of Predictors of Threat Effects (Posttest–Pretest)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictors</td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td>Arithmetic fluency</td>
</tr>
<tr>
<td>Processing speed</td>
</tr>
<tr>
<td>Group</td>
</tr>
<tr>
<td>Group × Arithmetic Fluency</td>
</tr>
<tr>
<td>Group × Processing Speed</td>
</tr>
</tbody>
</table>

---

**Figure 4.** Relations between threat and performance as a function of low ($−1 \text{ SD}$) and high ($+1 \text{ SD}$) levels of arithmetic fluency.
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 that has no (or much 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 nonstigmatized groups differ in baseline performance (see Jussim, Crawford, Anglin, Stevens, & Duarte, 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., Sack et al., 2003; Lemaire et al., 2018). Indeed, like Hess et al. who found deleterious effects of ABST may also occur in other domains where we often see much less age-related differences. These include moral reasoning (e.g., Margoni, Geipel, Hadjichristidis, & Surian, 2018; McNair, Okan, Hadjichristidis, & de Bruin, 2018), wisdom (e.g., Grossmann et al., 2010), and decision making (e.g., see Lückenhoff, 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 Experiment 1) or directly (computational estimation tasks in Experiments 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 (Experiment 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-verification strategy on false equations, when a quick parity-violation checking strategy was used in the control condition. It might seem strange at first that older adults under threat may choose a more demanding exhaustive calculation strategy over a quick parity-violation checking strategy. However, we can explain it in two ways. First the use of quick parity-violation checking
strategy requires to allocate attention resources to semantic features of the problems (Krueger, 1986; Krueger & Hallford, 1984) to detect that parity of proposed and correct answers differ, as well as cognitive flexibility to switch from a possibly default, exhaustive-verification strategy to a quick parity-violation checking strategy. It is likely that threat affected older adults’ available processing resources, impairing their analyses of semantic features of problems and their cognitive flexibility. As a consequence, it might seem 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 that could be resource demanding, especially under threat. Because the quick parity-violation checking strategy does not yield correct response to all problems, older adults under threat used the more resource-demanding, exhaustive-verification strategy on all problems.

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 strategy less often and/or to execute it more poorly. However, with the 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 (Experiments 2 and 3). One limitation of the present study is the relatively limited sample size to conduct mediation analyses. Moreover, the bootstrapping (pursued in our study) as a resampling method to construct reliable confidence does not resolve this issue (see, Koopman, Howe, Hollenbeck, & Sin, 2015, for a discussion). Therefore, results of mediation analyses from bootstrapping within a small sample like ours should be taken with caution, as there is always an inherent risk of false positives. Although our mediation model could have definitely used a larger sample size to confirm these findings, direct and indirect results of strategy examination from all three experiments provide converging evidence that threat effects occur via disrupting older adults’ strategy use. Thus, we can be confident that our mediation results are not spurious.

Interestingly, although strategy use mediated threat effects on both homogeneous and heterogeneous problems (Expt. 2), we found ABST effects on strategy use while participants solved homogeneous problems but not while solving heterogeneous problems. This may seem odd because heterogeneous problems are easier to solve than heterogeneous problems, and previous studies have shown that participants select the better strategy more often on homogeneous than on heterogeneous problems (Hodzik & Lemaire, 2011; Lemaire et al., 2004; LeFevre et al., 1993). Our results can be explained by comparing strategy performance of older adults under threat and control groups. In the control group, we replicated findings that participants perform better on homogeneous problems and are poorer in selecting the better strategy on heterogeneous problems. This suggests that the baseline performance was good on homogeneous but poor on heterogeneous problems. As a result, under ABST, when participants reduced their use of the better strategy, the difference in performance (control-threat) was larger on homogeneous than on heterogeneous problems.

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 (Experiments 2 and 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 suggest that ABST led older adults to be less efficient to inhibit the just executed strategy, to reactivate 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 & Lecacheur, 2010; Luwel, Onghena, Torbeys, Schillemans, & Verschaffel, 2009). This is also consistent with the hypothesis that stereotype threat impairs executive control mechanisms (Croizet et al., 2004; Mazerolle et al., 2012; Schmader & Johns, 2003; Schmader et al., 2008; Schmader & Beilock, 2012).

Interestingly, we found reduced versus exacerbated threat effects on different measures (i.e., speed and accuracy) of complex arithmetic performance (Experiments 2 and 3). One distinct pattern was an increased speed in older adults under threat leading them to often be as fast as older adults in control group (e.g., on heterogeneous problems in Experiment 2 and on target problems in Experiment 3), and sometimes even faster (e.g., on homogeneous problems in Experiment 2 and on priming problems in Experiment 3). This increased speed was accompanied by dramatically less precise estimates. This means that the threat manipulation increased participants’ speed and decreased the precision of their estimates. These findings can again be explained as a result of change in strategy choice under threat. Being less careful at selecting the better strategy and further repeating this poor strategy choice on consecutive trials led older adults under threat to be faster but less accurate, even if task instructions required them to try their best to provide most accurate estimates. A limitation of this study was that we did not investigate how participants made strategy choice decisions, and in particular if these decisions were the same or different under threat and control conditions, an issue that future studies may address by probing participants explicitly (via verbal protocols e.g.).

Another finding consistent with the hypothesis that ABST decreased older adults’ executive control processes concerns participants’ performance on trials where strategy selection was controlled (Experiment 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 cue 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; Desrichard & 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, 2013a, 2013b). 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 & Förster, 2004). It is likely that in our experiments ABST led older adults to adopt a more prevention focus (i.e., increased vigilance toward 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 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 et al., 2006; Abrams et al., 2008), or higher levels of dementia-worry (e.g., Fernández-Ballesteros, Bustillos, & Huici, 2015; Fresson, Dardenne, Geurten, & Meulemans, 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, 2009).

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 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 Psychosocial 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, Petitit, Little, & Chiu, 2008; Fratiglioni, Paillard-Borg, & Winblad, 2004), social support (Bélard, Zanunegui, Alvarado, Otero, & Del Ser, 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 (Owby, Crocco, Acevedo, John, & Loewenstein, 2006; Royall, Palmer, Chiordo, & Polk, 2012), subjective feelings of loneliness (Holwerda et al., 2014), individual personality traits (Caselli et al., 2016; Kuzma, Sattler, Toro, Schönknecht, & Schröder, 2011; Luchetti, Terracciano, Stephan, & Sutin, 2016; 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 article are in the same research institution. They had plenty of opportunities to exchange about their work and research agenda, Patrick Lemaire on strategic aspects of mathematical cognition and on cognitive aging and Isabelle Régner on effects of stereotype threat. When Patrick Lemaire and Poshita Nicolas heard Isabelle Régner 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 Isabelle Régner and Poshita Nicolas read and heard Patrick Lemaire’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


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


AGE-BASED STEREOTYPE THREAT


This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.


Steele, C. L., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. Advances...


Members of Underrepresented Groups:
Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write APA Journals at Reviewers@apa.org. Please note the following important points:

• To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.

• To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.

• To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In the letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, “social psychology” is not sufficient—you would need to specify “social cognition” or “attitude change” as well.

• Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.

APA now has an online video course that provides guidance in reviewing manuscripts. To learn more about the course and to access the video, visit http://www.apa.org/pubs/journals/resources/review-manuscript-ce-video.aspx.