Memory & Cognition
https://doi.org/10.3758/s13421-023-01488-2

Overcoming age differences in memory retrieval by reducing stereotype threat

Marie Mazerolle[1] · Lucas Rotolo[1] · François Maquestiaux[1,2]

Accepted: 27 October 2023

Abstract
Very little is known about whether and how socioemotional factors influence age differences in associative memory. Here, we tested the hypothesis that reducing the threat induced by age-based stereotypes can reduce age differences in learning performance and strategy. Using an associative learning task, we replicated the classic finding of age differences under a high-threat condition: older adults had longer reaction times than younger adults and were much more reluctant to use memory retrieval. However, age differences were greatly diminished under a low-threat condition. These findings demonstrate that memory retrieval is an ability not entirely lost as individuals age because merely reducing stereotype threat helped restoring it. We conclude that socioemotional factors, such as stereotype threat, should be considered when evaluating younger and older adults' memory performance.

Keywords Cognitive aging · Stereotype threat · Memory · Associative learning · Cognitive performance

Introduction

Memory is a central function of everyday life and the cause of numerous complaints in the general population and in older adults in particular (Jonker et al., 2000). Indeed, memory is one of the first cognitive functions affected by aging (Zacks et al., 2000), especially when older adults have to memorize associative information (Naveh-Benjamin, 2000). Recent studies have challenged the habitual deficit-oriented interpretation of age differences in associative memory by highlighting the crucial importance of socioemotional factors and strategic approaches to tasks (Touron, 2015). Here, we manipulated stereotype threat, a powerful socioemotional factor known to significantly decrease older adults' cognitive performance (Barber, 2017), to examine its precise influence on the magnitude of age differences in associative learning. Importantly, we hypothesized that reducing stereotype threat can greatly diminish age differences in memory and learning performance and strategy.

The cognitive aging literature has established that older adults experience particular difficulty memorizing relationship information about items, a phenomenon termed *associative deficit* (Naveh-Benjamin, 2000). Structural and functional imaging studies of associative deficits have revealed the key role played by the medial temporal lobe, the posterior parietal cortex, and the episodic network (Duarte & Dulas, 2020). However, some studies have challenged this deficit perspective by showing some malleability in age-related associative deficits. For example, when relying on schematic support or existing knowledge, younger and older adults can produce similar performance (Castel, 2005). Older adults' associative deficit also appears to be more related to strategic choices than to cognitive deficits per se (Touron, 2015). This hypothesis was derived from studies using the noun-pair lookup task (Ackerman & Woltz, 1994). The goal of this task is to decide whether a pair of nouns (the target) matches or does not match one of the pairs displayed in a table located at the top of a computer screen. Two main strategies can be used to perform the noun-pair lookup task: visual scanning of the table or direct memory retrieval. Early in practice, participants mainly rely on visual scanning, a strategy producing long reaction times. And as practice increases, participants tend to learn the pairs more and more, and so become more and more able to use memory

✉ Marie Mazerolle
marie.mazerolle@univ-fcomte.fr

[1] Université de Franche-Comté, LINC, MSHE, F-25000 Besançon, France

[2] Univ Rouen Normandie, Normandie Univ, CRFDP UR 7475, F-76000 Rouen, France

retrieval, a strategy producing short reaction times. Thus, the noun-pair lookup task offers two converging behavioral indicators that allow researchers to assess associative learning performance: an objective measure (reaction time, RT) and a self-reported measure of the strategy used (i.e., visual scanning vs. memory retrieval). Touron and Hertzog (2004a) found that older adults were more reluctant than younger adults to adopt memory retrieval, as evidenced by longer RTs and less reports of memory retrieval. However, in conditions promoting the use of memory retrieval (e.g., when some trials present a target pair without the table, not allowing participants to use scanning), age-related differences in learning and strategy use were greatly reduced. Importantly, participants' confidence in their memory influenced their use of memory retrieval: older adults with the least confidence in their memory tended to avoid retrieval more when performing the associative task. This study elegantly demonstrated that associative deficits encountered with advancing age might be more malleable than previously thought. Following Touron and Hertzog (2004a), other studies have identified other socioemotional factors influencing the effects of aging on cognition, such as motivation (Zhang et al., 2013), effort (Maquestiaux & Ruthruff, 2021), or stereotype threat (Barber, 2017).

Stereotype threat can be defined as the fear that stigmatized people feel when they are aware their performance could confirm negative stereotypes about their social group (Steele & Aronson, 1995). In addition to the normal anxiety associated with taking cognitive tests, stereotype threat creates an extra pressure that interferes with stigmatized individuals' cognitive functioning (Spencer et al., 2016). Because older adults are often stereotyped as being senile and less cognitively capable (Hummert, 2011), they can experience stereotype threat when undergoing cognitive evaluation and show reduced performance on various physical and cognitive tests (Lamont et al., 2015). Episodic memory appears particularly affected by stereotype threat, which reduces the amount of attention that can be used to control memory processes (Mazerolle et al., 2012) and implement retrieval strategies (Hess et al., 2003). Importantly, reducing stereotype threat can greatly diminish, and sometimes completely eliminate, age differences in memory performance in these studies. Brubaker and Naveh-Benjamin (2018) demonstrated that stereotype threat significantly impaired older adults' associative performance; however, the exact mechanism that underlies these effects was not precisely evaluated.

The current study

Here, we evaluated whether age differences usually observed in associative memory might reflect the influence of socioemotional factors, and especially stereotype threat. The key manipulation of the present study was to increase or reduce the activation of stereotype threat in younger and older adults performing an associative learning task. We used the noun-pair lookup task to test the hypothesis that reducing stereotype threat should promote the use of memory retrieval, thereby facilitating older adults' performance and reducing age effects in associative memory.

One important issue, as yet unresolved, relates to the mechanisms underlying the impact of stereotype threat on younger and older adults' performance. One of the most influential frameworks for understanding the impact of stereotype threat on stigmatized individuals' performance is the executive resources depletion (ERD) hypothesis (Schmader et al., 2008). Following this account, stereotype threat triggers physiological stress responses (Blascovich et al., 2001), suppression processes devoted to the regulation of negative thoughts and emotions (Cadinu et al., 2005), and performance of self-monitoring processes (Beilock & DeCaro, 2007). Together, these processes and responses disrupt executive processing (Schmader & Johns, 2003), thus reducing executive resources available to perform the task.

An alternative hypothesis posits that stereotype threat leads to motivational changes, influencing stigmatized individuals' approach of the task. Based on Higgin's (1997) regulatory focus framework, Seibt and Förster (2004) demonstrated that under stereotype threat, individuals adopt a prevention focus, increasing their vigilance to avoid making errors. This motivational change results in a more cautious approach of the task, characterized by reduced speed but increased accuracy.

The ERD and prevention focus hypotheses have received empirical support in both younger adults (Schmader & Johns, 2003; Seibt & Förster, 2004) and older adults (Barber & Mather, 2013; Mazerolle et al., 2012). However, there have been few attempts to evaluate them within a single study comparing the impact of stereotype threat in both younger and older adults. In one of them, Popham and Hess (2015) used a letter-canceling task to examine the prevention focus hypothesis and an operation span task to examine the ERD hypothesis. Stereotype threat reduced speed and increased accuracy in both younger and older adults on the letter-canceling task, suggesting the adoption of a prevention focus. Importantly, in line with the ERD hypothesis, stereotype threat only affected younger adults' working-memory performance. Although these results might suggest that different mechanisms underlie stereotype threat effects in younger and older adults, it should be noted that the operation span task was not explicitly termed a memory test "in order to minimize potential threat effects in older adults" (Popham & Hess, 2015, p. 226). It therefore seems likely that this way of presenting the task might have prevented stereotype threat from occurring in older adults, making it difficult, if not impossible, to evaluate the ERD hypothesis.

In the present study, we examine the impact of stereotype threat in younger and older adults using a single task, the noun-pair lookup task, therefore making it possible to directly test the predictions of the ERD and prevention focus hypotheses. According to prevention focus hypothesis, stereotype threat should cause participants to adopt a cautious approach when completing the task. Concretely, stereotype threat should cause a *retrieval avoidance* (i.e., longer RTs and less reports of memory retrieval as compared to younger adults), as visual scanning is a longer but safer strategy for responding. ERD also predicts a *retrieval avoidance* under stereotype threat as performance should especially suffer when controlled resources are needed and memory retrieval requires cognitive resources (at least early in practice). However, it also allows us to make a specific prediction about RTs as a function of the strategy used. Indeed, executive depletion should result in a specific slowdown on trials for which participants use memory retrieval but not visual scanning, as the former strategy relies more on executive resources. This prediction is unique to the ERD hypothesis, as the hypothesis of prevention focus would predict a general slowdown, regardless of the cognitive cost of the strategy (Barber, 2017).

Method

Participants

Forty younger adults and 40 older adults were included in the study sample. Participants from each age group were randomly assigned to either a high-threat condition ($n = 20$) or a low-threat condition ($n = 20$). The sample size of 40 participants per age group was fixed in advance and chosen so that we would have as much statistical power as the study by Touron and Hertzog (2004a), which inspired the present work. Indeed, Touron and Hertzog (2004a) had 40 younger and 40 older adults performing an associative task (roughly similar to the one used here) with age group and memory-probe (no-memory vs. memory probes) as between-subject factors and practice blocks as a within-subject variable, a design roughly similar to the age group × condition × practice block design of the present study.

The younger adults ($n = 40$; $M_{age} = 20$ years; $SD_{age} = 1.7$ years; range: 17–27 years; 32 women) were undergraduate psychology students recruited from psychology courses at the University of Franche-Comté. The older adults ($n = 40$; $M_{age} = 67.2$ years; $SD_{age} = 6.1$ years; range: 59–88; 17 women) were primarily recruited by word of mouth (e.g., family friends, community groups) or from existing university participant pools. All were native French speakers. Before being included in the sample, participants were prescreened for psychological and neurological health issues and could not be taking any medications that might affect cognition (e.g., antidepressants, drugs with anticholinergic properties, benzodiazepines, opiates, and/or anticonvulsants). There was no significant difference in the number of years of education between younger and older participants ($M_{age\ young} = 13.5$ years; $SD = 0.8$; vs. $M_{age\ old} = 14.08$ years; $SD = .3.6$), $t(79) < 1$. In line with the literature, older adults demonstrated higher vocabulary abilities than younger adults ($M_{young} = 26$; $SD = 4.5$ vs. $M_{old} = 19.6$; $SD = .5.6$) on the Mill Hill vocabulary test (Deltour, 1993), $t(79) = 5.63$, $p < .001$.

Older participants were additionally screened for cognitive impairment at the end of the experiment, using the Mini-Mental State Examination (Folstein et al., 1975), and reached or exceeded a cutoff score corresponding to their age and educational level (Crum et al., 1993).

Materials

We developed a noun-pair lookup task (see Fig. 1) that was inspired by Touron and Hertzog (2004a) and implemented using E-Prime software (Version 2.1; Psychology Software Tools, Pittsburgh, PA, USA).

In each noun-pair lookup task trial, a target item consisting of two nouns (e.g., *doll – wheel*) was presented in the center of the screen, while a table of six noun pairs was always presented at the top of the screen. The words used for the noun pairs were selected from a list of French words taken from Bonin et al. (2003). Noun pairs did not differ in values of concreteness, imagery, frequency, and emotional valence (Fs < 1), and the associative value of words was controlled between pairs of words based on Ferrand and Alario's (1998) French norms.

The location of the noun pairs in the table varied randomly from one trial to the next. Half of the trials presented a target item that matched one of the pairs shown in the table (e.g., *doll – wheel*), while the other half presented a target item that did not match (e.g., *doll – storm*). These non-matching trials paired a left-hand word from one pair of the table with a randomly selected right-hand word from a different pair of the table.

For each trial, participants indicated on an AZERTY keyboard whether (press the *K* key for "Yes") or not (press the *L* key for "No") the target pair matched a pair in the table. The Yes/No responses were given with the index and middle finger of the right hand. The participants were then asked to indicate the strategy they had used to answer by pressing a key with their left hand: visual scanning (press "V", *S* key), memory retrieval (press "M", *D* key), both strategies (press "2", *F* key), another strategy (press "A", *G* key).

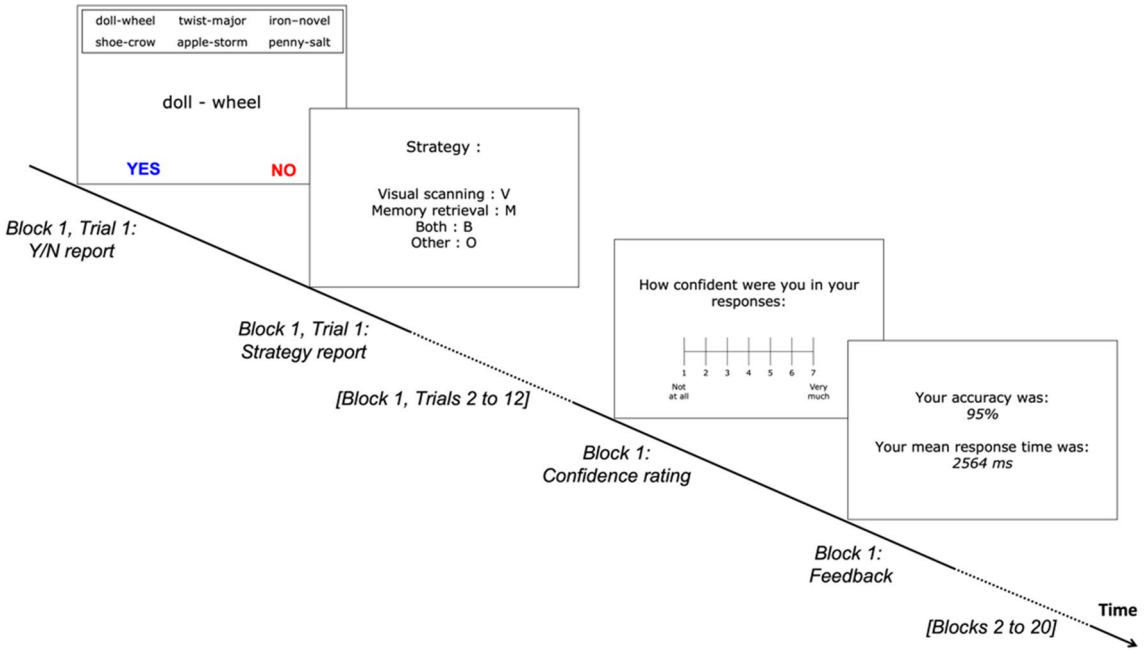

Fig. 1 Overview of the noun-pair lookup task

Procedure

After giving their informed consent, younger and older participants were randomly assigned to either a high-threat condition or a low-threat condition. They first completed a familiarization block of the noun-pair lookup task, with similar materials and procedure to the 20 blocks they would perform later. After familiarization, older and younger participants performed the task depending on the threat condition they were assigned to. In the high-threat condition, older adults were told that that they were going to perform a memory task and that both younger and older adults were taking part in the study. Additionally, the experimenter asked participants their age and typed it on the keyboard, so that it appeared on the screen. In the low-threat condition, older adults received the same instruction as in the high-threat condition, but they were also told that there was typically no age difference on this type of task (i.e., that the task was age-fair). The task was also presented in this way to the half of the younger adult sample assigned to the low-threat condition. We assumed that nullifying age-related comparisons would reduce stereotype threat in older adults and not induce any threat in younger adults, as had previously been observed by Mazerolle et al. (2012). The remaining half of the younger adult sample was assigned to a high-threat condition. These participants were told that they were going to perform a task that measured intellectual ability and that both psychology and medical students were taking part in the study. Additionally, the experimenter asked them their academic field and typed it on the keyboard, so that it appeared on the screen. We assumed that comparing psychology majors to medical majors would elicit stereotype threat, as had previously been observed by Dutrévis and Croizet (2005).

Immediately after the induction phase (high-threat or low-threat), participants performed the noun-pair lookup task. There were 20 repetitions of each of the 12 items, resulting in 20 blocks of 12 trials, each consisting of six matched and six non-matched trials. At the end of each block, participants rated how confident they were about the accuracy of their responses on a scale ranging from 1 to 7. Feedback on the accuracy and mean RT on the block was then displayed on the screen. Participants were instructed to take longer to respond if their accuracy was below 95% or to speed up if it was above 95%. The session lasted approximately 1 h.

Results

Reaction times

We removed trials for which RTs were below 300 ms (0.03%) or above 3 standard deviations (SDs) from the mean of each age group (younger adults: 1.03%; older adults: 1.01%). Error trials were also removed from RT analyses.[1]

[1] Overall accuracy was close to the instructed 95% level, but slightly lower for older than for younger adults (.92 vs. .96), $F(1, 76) = 34.12$, $p < .001$, $\eta_p^2 = .31$. Threat condition and practice did not influence accuracy, nor interacted with age group.

Figure 2 shows mean RT (top panel) and the mean proportion of retrieval strategy reports (bottom panel) as a function of practice blocks (1–20) for each of the four groups (2 age groups × 2 threat conditions).

A mixed ANOVA was conducted on mean RT for correct responses, with age group (younger and older adults) and condition (high-threat and low-threat) as between-subjects factors and practice block as a within-subjects factor. The main effect of practice block was significant, $F(19, 58) = 11.61$, $p < .001$, $\eta_p^2 = .79$, with a gradual RT shortening from block to block (from 2,978 ms in block 1 to 2,107 ms in block 20). Younger adults were overall faster than older adults (2,111 ms vs. 2,609 ms over the 20 blocks), $F(1, 76) = 28.59$, $p < .001$, $\eta_p^2 = .27$. These main effects were considered in the context of a significant age × practice interaction, $F(19, 58) = 2.87$, $p = .001$, $\eta_p^2 = .48$, indicating that the shortening of RT with practice was more pronounced for younger than for older adults (reduction of 947 vs. 794 ms over the 20 blocks). The main effect of condition was not significant ($F(1, 76) < 1$), but interacted with age group, $F(1, 76) = 12.86$, $p < .001$, $\eta_p^2 = .14$. Planned comparisons in each age group revealed that the high-threat condition lengthened older adults' RTs by 302 ms as compared to the low-threat condition (2,760 vs. 2,458 ms over the 20 blocks, respectively), $F(1,76) = 5.26$, $p = .025$, $\eta_p^2 = .06$. However, the reverse pattern was observed in younger adults, with RTs being 336 ms shorter in the high-threat condition than in the low-threat condition (1,928 vs. 2,294 ms over the 20 blocks respectively), $F(1,76) = 7.72$, $p = .007$, $\eta_p^2 = .09$. Importantly, younger adults responded much faster than older adults (1,928 vs. 2,760 ms over the 20 blocks) in the high-threat condition, $F(1, 76) = 39.902$, $p < .001$, $\eta_p^2 = .34$, whereas age differences were not significant in the low-threat condition (2,294 vs. 2,458 ms over the 20 blocks), $F(1,76) = 1.55$, $p = .217$.

Retrieval strategy reports

A mixed ANOVA was conducted on the mean proportion of reported retrieval strategy, with age group (younger and older adults) and condition (high-threat and low-threat) as between-subjects factors and practice block (1–20) as a within-subjects factor. A main effect of practice indicated that the proportion of memory retrieval increased with repetition (from .10 in Block 1 to .50 in Block 20), $F(19, 56) = 4.84$, $p < .001$, $\eta_p^2 = .62$. Younger adults used memory retrieval more often than older adults (.47 vs. .28 over the 20 blocks), $F(1, 74) = 8.54$, $p = .005$, $\eta_p^2 = .10$. Importantly, the age × condition interaction was significant, $F(1, 74) = 11.18$, $p = .001$, $\eta_p^2 = .13$. Planned comparisons in each age group indicated that the low-threat condition increased older adults' use of memory retrieval by 22% as compared to the high-threat condition (.39 vs. .17 over the 20 blocks, respectively), $F(1, 74) = 6.11$, $p = .016$, $\eta_p^2 = .08$. Younger adults exhibited the opposite pattern, using memory retrieval 20% more in the high-threat condition than in the low-threat condition (.57 vs. .37 over the 20 blocks, respectively), $F(1, 74) = 5.07$, $p = .027$, $\eta_p^2 = .06$. Importantly, younger adults reported using retrieval much more frequently than older adults (.57 vs. .17 over the 20 blocks) in the high-threat condition ($F(1, 74) = 19.63$, $p < .001$, $\eta_p^2 = .21$), whereas age did not significantly influence the use of memory retrieval in the low-threat condition ($M_{young} = .37$ vs. $M_{old} = .39$ over the 20 blocks ; $F < 1$).

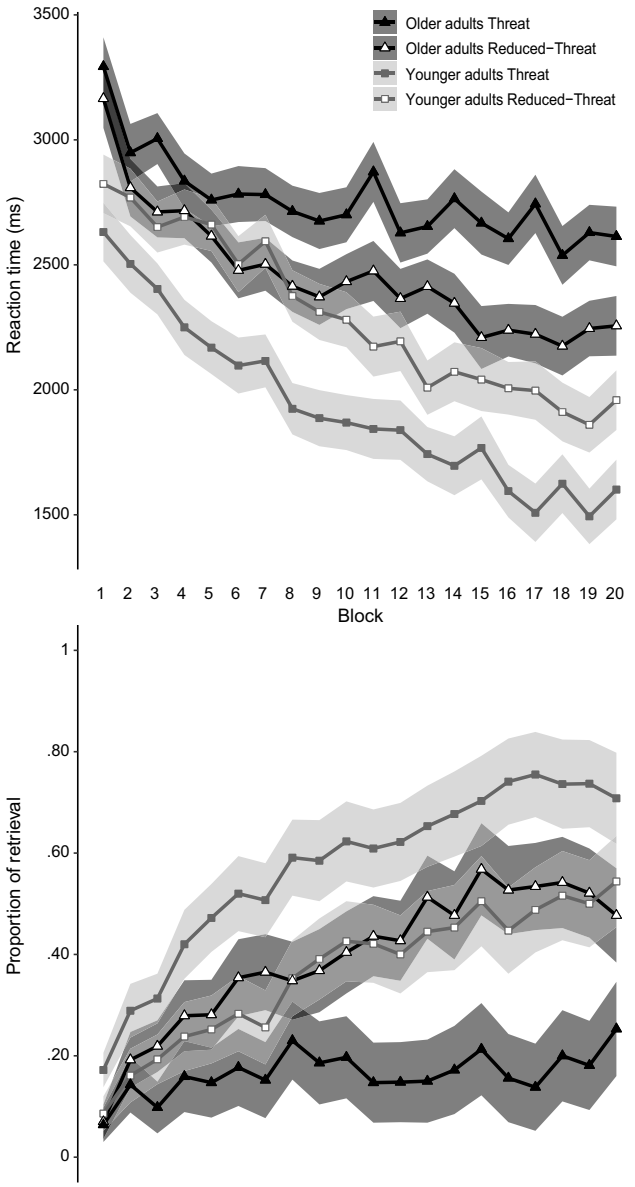

Fig. 2 Mean reaction time (**top panel**) and mean proportion of retrieval strategy reports (**bottom panel**) as a function of practice blocks, age group (Younger adults vs. Older adults), and condition (High-threat vs. Low-threat). The shaded areas around the lines represent standard error of the mean

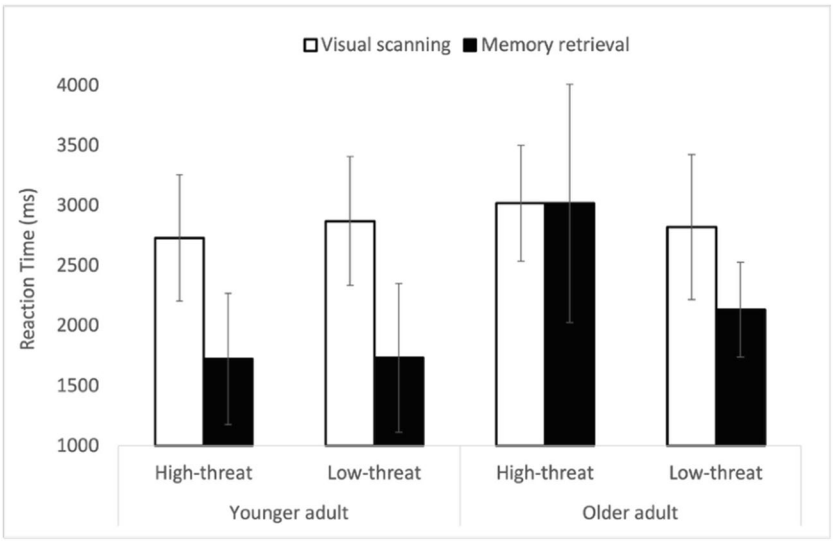

Fig. 3 Mean reaction time as a function of reported strategy (scanning vs. retrieval), age (younger adult vs. older adult), and condition (high-threat vs. low-threat). Error bars indicate standard error of the mean

Reaction times as a function of reported strategy

Reports of *both* (i.e., *scanning* and *retrieval*) and *other* strategies were quite rare, occurring for approximately 11% and 0% of the responses, respectively. Older adults reported using *both* strategies (M = .03) less often than younger adults (M = .17), $F(1, 76) = 15.38$, $p < .001$, $\eta_p^2 = .17$. Threat condition neither influenced strategy reports nor interacted with age effects (Fs <1). Because of the very low percentages of *other* and *both* strategy reports, RTs were compared only for the *scanning* (52%) and *retrieval* (37%) strategies. Figure 3 shows the resulting mean RTs as a function of reported strategy (scanning vs. retrieval) for each of the four groups (2 age groups × 2 threat conditions).

An ANOVA was conducted on mean RT for correct responses, with age group (younger and older adults), condition (high-threat and low-threat), and strategy (scanning vs. retrieval) as between-subjects factors. A main effect of strategy revealed that RTs were faster by 688 ms on trials with reported memory retrieval (M = 2,112 ms) than on trials with reported visual scanning (M = 2,800 ms), $F(1, 61) = 82.25$, $p < .001$, $\eta_p^2 = .57$. Additionally, older participants' RTs were slower than those of younger adults (2,674 vs 2,239 ms), $F(1, 61) = 16.89$, $p < .001$, $\eta_p^2 = .22$. These main effects were considered in the context of a significant strategy × age group interaction, $F(1, 61) = 16.81$, $p < .001$, $\eta_p^2 = .22$. Planned comparisons indicated that older adults' RTs were slower than those of younger adults on trials with reported memory retrieval (2,485 ms vs. 1,738 ms), $F(1, 61) = 27.59$, $p < .001$, $\eta_p^2 = .31$, whereas age did not significantly influence RTs on trials with reported visual scanning (2,863 ms vs. 2,738 ms), $F(1, 61) = 1.12$, $p = .294$. The condition × age group interaction was also significant, $F(1, 61) = 9.21$, $p <= .004$, $\eta_p^2 = .13$, with older adults being slower than younger participants by 757 ms in the high-threat condition (2,905 vs. 2,148 ms, respectively), $F(1, 61) = 24.09$, $p < .001$, $\eta_p^2 = .28$, but not in the low-threat condition (a nonsignificant difference of 114 ms, 2,443 vs. 2,329 ms), F <1. Importantly, the three-way age × condition × strategy interaction was significant, $F(1, 61) = 7.88$, $p = .007$, $\eta_p^2 = .11$. Planned comparisons revealed that stereotype threat slowed older adults' RTs by 761 ms on trials with reported memory retrieval as compared to the low-threat condition (2,866 vs. 2,105 ms), $F(1, 61) = 13.20$, $p <.001$, $\eta_p^2 = .18$, whereas this effect was not significant on trials with reported visual scanning (2,944 vs. 2,782 ms), F <1. Younger and older adults' RTs differed on trials with reported memory retrieval in the high-threat condition (1,585 vs. 2,866 ms), $F(1, 61) = 38.35$, $p < .001$, $\eta_p^2 = .38$, but not in the low-threat condition (1,893 vs. 2,104 ms), $F(1, 61) = 1.18$, $p =.281$. There was no significant age difference on trials with reported visual scanning in either the high-threat condition ($F(1, 61) = 1.85$, $p =.178$) or the low-threat condition (F <1).

Confidence

Confidence judgments were averaged across blocks, and data from three participants (two younger adults and one older adult) were excluded as their responses were not fully recorded. The main effect of condition was significant ($F(1, 73) = 3.98$, $p = .050$, $\eta_p^2 = .05$), with participants in the high-threat conditions (M = 6.33) being slightly less confident than those in the low-threat conditions (M = 6.63).

Neither the main effect of age nor the age × condition interaction was significant.

Discussion

The present study examined whether age differences in associative memory are influenced by socioemotional factors. We predicted that older adults' tendency to avoid direct memory retrieval while learning new associations could be reduced by deactivating age-based stereotype threat. Using the noun-pair lookup task, we replicated classic age differences among participants assigned to a high-threat condition: relative to younger adults, older adults were slower and used memory retrieval less. However, when younger and older adults were assigned to a low-threat condition, age differences vanished. Taken together, these results question the origin of the age-related differences usually observed in associative learning and suggest that memory retrieval is not lost but is instead a capacity that can be regained when stereotype threat is reduced. In line with studies that have observed the prejudicial impact of stereotype threat on older adults' cognitive performance (Armstrong et al., 2017) and increased age-related differences under stereotype threat (e.g., Mazerolle et al., 2012), this study suggests paying special attention to the influence of stereotype threat when evaluating older adults' memory.

The results also provide insights about the mechanisms that underlie the impact of stereotype threat on younger and older adults' cognitive performance. Even though stereotype threat reduced confidence in both younger and older adults, it seems to impact younger and older adults' associative learning in opposite ways. In older adults, stereotype threat slowed RTs and prevented them from switching from visual scanning (a longer but safer strategy) to memory retrieval (an effective but cognitively more demanding strategy early in practice). This *retrieval avoidance* nicely fits with the prevention focus hypothesis of stereotype threat (Seibt & Förster, 2004), which predicts a cautious approach of the task. Importantly, stereotype threat also seemed to prevent older adults from automatizing their memory retrieval processes, as demonstrated by the specific RT lengthening on trials with reported memory retrieval. This result is consistent with the executive resources depletion hypothesis (ERD, Schmader et al., 2008), which predicts performance impairment when cognitively demanding processes are needed. Taken together, these results are consistent with both the ERD and prevention focus theoretical accounts, and suggest that stereotype threat likely induced a prevention focus but also depletes executive resources in older adults. Indeed, we argue that ERD and prevention focus likely interact, a suggestion already made by Mazerolle et al. (2021). That is, we propose that the vigilant, risk-averse processing style caused by a prevention focus adoption (Seibt & Förster, 2004), and the executive processing disruption caused by monitoring processes to analyze self-performance (Schmader et al., 2008), likely document the same phenomenon: an increased vigilance toward errors and signs of failures under stereotype threat that slows responding and disrupts performance when using the most costly cognitive strategies. Despite the current results appearing to be consistent with both ERD and prevention focus accounts of stereotype threat, further research is needed to better understand the interaction between the two hypotheses.

In younger adults, stereotype threat resulted in the opposite pattern to that of older adults: stereotype threat shortened RTs and increased memory retrieval. Although improvements in associative learning performance and strategy under stereotype threat may appear surprising at first sight, this result is broadly consistent with previous studies that have demonstrated improved performance on easy tests (O'Brien & Crandall, 2003) and can be interpreted in the light of the mere effort hypothesis (Harkins, 2006). According to the mere effort hypothesis, stereotype threat motivates individuals to perform well (supposedly to disconfirm the stereotype), and thus potentiates the dominant response on the task. Because the dominant response is generally correct in easy tasks, the extra motivation generated by stereotype threat tends to produce better performance (Ben-Zeev et al. 2005). In line with social facilitation findings, when people fear being evaluated, their dominant responses are strengthened (Zajonc, 1965). The mere effort hypothesis seems particularly relevant for interpreting younger adults' response patterns since their accuracy was 96%, and the RT improvement in the high-threat condition was not specific to trials with reported memory retrieval (as is the case with older adults), but instead reflected a general RT reduction.

The present research allowed us to directly compare the impact of stereotype threat in younger and older adults. Despite stereotype threat lowering confidence in both younger and older adults, it increased younger adults' performance but weakened older adults' performance. However, this pattern might be interpreted as evidence in favor of different mechanisms underlying stereotype threat in younger and older adults. As proposed by Popham and Hess (2015), we argue that task difficulty can also play a key role. In our study, participants only had to memorize six word-pairs, a design previously labeled as a "low memory load" condition that clearly appears as a condition that promotes age differences on RTs early in practice (Touron and Hertzog, 2004b). That is, memorizing six pairs of words seemed to be far more difficult for older adults in Touron and Hertzog (2004b), and also in our study (at least in the high-threat condition), as supported by accuracy differences in younger and older adults in favor of younger ones. Because task difficulty is a clear moderator of stereotype threat effects both in younger

(O'Brien & Crandall, 2003) and in older adults (Hess et al., 2009), we argue that our design might be responsible for the different mechanisms that seem to underlie stereotype threat effects in the two age groups: performance improvement on an easy task in younger adults, and performance decrement on a difficult memory test in older adults.

Limitations and future directions

Our study allows direct comparison of the impact of stereotype threat in younger and older adults, but we had to use different stereotype inductions in the high-threat condition: instructions that targeted aging stereotypes for older adults, whereas instructions that targeted academic performance for younger adults. Despite these differences, both inductions proved effective in inducing stereotype threat both in older (e.g., Mazerolle et al., 2012) and in younger adults (e.g., Dutrévis & Croizet, 2005); however, we cannot rule out that differences in these two stereotypes might have differentially influenced the impact of stereotype threat, thereby limiting comparison between the two age groups. Along the same line, the low-threat condition directly negates the impact of age on memory performance but not the stereotypes of academic performance. Despite this low-threat condition being clearly less threatening than our high-threat condition (as previously demonstrated in Mazerolle et al., 2012, 2015), there is a possibility that younger participants in the low-threat condition might still have been threatened (as well as older adults, who might have also been threatened beyond aging stereotypes by some characteristics of the situation, such as evaluative pressure, or the presence of a younger experimenter). Future studies should use the exact same stereotype in younger and older adults to allow more direct comparison, for example, using math stereotypes and testing younger and older women.

It is noteworthy that our design (six word-pairs across 20 repetitions) might also have altered participants' transition from visual scanning to memory retrieval, as previously observed in Touron and Hertzog (2004b). Indeed, in our study after 20 repetitions, younger adults used retrieval about 70% of the time in the high-threat condition versus 50% in the low-threat condition, whereas older adults used it about 50% of the time in the low-threat condition and 20% in the high-threat condition. In comparison, using set size of 20 word-pairs usually allows younger adults to retrieve about 80% of the time and 40% of the time in older adults after 20 repetitions (Touron, 2015). In our study, using a small set-size might have made visual scanning too easy, and efficient enough not to need switching to retrieval. In line with this speculation, Hertzog et al., (2007) observed that transition to retrieval was driven by participants' estimated RT benefit of retrieval over scanning. In our study, it seems likely that younger adults in the low-threat condition were not motivated enough to use memory retrieval as they might assume it would not improve their RTs that much. However, in the high-threat condition, the extra motivation generated by stereotype threat seems to have improved their RTs and drove them to use memory retrieval more (to the levels usually observed in the literature with larger set sizes). In older adults, our results demonstrated that stereotype threat decreased the use of memory retrieval, likely because older adults noticed that their RTs did not decrease when using it. However, under a low-threat condition, older adults' RTs were shortened when using memory retrieval, leading them to use it more (and beyond the levels usually observed in the literature with larger set sizes). Future studies should be conducted with a larger set size and much more repetition, to promote transition from visual scanning to memory retrieval and thus to permit automatization. Increasing set size and number of repetitions would indeed allow us to test the facilitatory effect of stereotype threat on automatized memory processes, as observed in younger adults and by Mazerolle et al. (2012) in older adults.

A key finding of the present study is that very subtle inductions of stereotype threat (i.e., emphasizing the memory component of the test and mentioning the participation of younger adults) greatly altered older adults' performance. This point deserves particular attention, since a typical mistake by newcomers to stereotype threat research is to assume that a stereotype threat condition requires the application of additional pressures that do not exist in conventional real-life testing, and that standard real-life testing instructions can be used to operationalize a no-stereotype threat control condition (Steele & Davies, 2003). In fact, the opposite is true: standard real-life testing settings are likely to induce implicit or explicit stereotype threat due to the words used to present the study or the tests (e.g., mentioning an "aging and memory study") or any environmental cues related to negative aging stereotypes (e.g., posters on Alzheimer's disease on the research laboratory walls). Here, simply mentioning or nullifying age-related comparison on a memory test was sufficient to greatly influence the magnitude of age differences on learning performance and strategy.

Open practices statement The study reported in this article was not preregistered. The data are available at https://lecloud-mshe.univ-fcomte.fr/index.php/s/W3gdE3YLmEjdaWq. Requests for the materials used in the study can be sent to the corresponding author.

Funding This research was supported by a grant from the Institut Universitaire de France awarded to François Maquestiaux.

Declarations

Ethics approval All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964

Helsinki Declaration and its later amendments or comparable ethical standards. The University of Franche-Comté Ethics Committee has confirmed that no ethical approval is required.

Competing interests The authors have no competing interests to declare that are relevant to the content of this article.

References

Ackerman, P. L., & Woltz, D. J. (1994). Determinants of learning and performance in an associative memory/substitution task: Task constraints, individual differences, volition, and motivation. *Journal of Educational Psychology, 86*(4), 487–515. https://doi.org/10.1037/0022-0663.86.4.487

Armstrong, B., Gallant, S. N., Li, L., Patel, K., & Wong, B. I. (2017). Stereotype threat effects on older adults' episodic and working memory: A meta-analysis. *The Gerontologist, 57*(suppl_2), S193–S205. https://doi.org/10.1093/GERONT/GNX056

Barber, S. J. (2017). An examination of age-based stereotype threat about cognitive decline. *Perspectives on Psychological Science, 12*(1), 62–90. https://doi.org/10.1177/1745691616656345

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

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

Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology, 41*(2), 174–181. https://doi.org/10.1016/J.JESP.2003.11.007

Blascovich, J., Spencer, S. J., Quinn, D., & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science, 12*(3), 225–229. https://doi.org/10.1111/1467-9280.00340

Bonin, P., Méot, A., Aubert, L.-F., Malardier, N., Niedenthal, P. M., & Capelle-Toczek, M.-C. (2003). Normes de concrétude, de valeur d'imagerie, de fréquence subjective et de valence émotionnelle pour 866 mots. *L'Année Psychologique, 103*(4), 655–694. https://doi.org/10.3406/PSY.2003.29658

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

Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? *Psychological Science, 16*(7), 572–578. https://doi.org/10.1111/J.0956-7976.2005.01577.X

Castel, A. D. (2005). Memory for grocery prices in younger and older adults: The role of schematic support. *Psychology and Aging, 20*(4), 718–721. https://doi.org/10.1037/0882-7974.20.4.718

Crum, R. M., Anthony, J. C., Bassett, S. S., & Folstein, M. F. (1993). Population-based norms for the Mini-Mental State Examination by age and educational level. *JAMA*, *269*(18), 2386–2391. https://doi.org/10.1001/JAMA.1993.03500180078038

Deltour, J. J. (1993). Echelle de vocabulaire de Mill Hill de JC Raven. *Adaptation française et normes européennes du Mill Hill et du Standard Progressive Matrices de Raven (PM38).* Braine-le-Château: Editions l'application des techniques modernes.

Duarte, A., & Dulas, M. R. (2020). Episodic memory decline in aging. In *The Cambridge handbook of cognitive aging: A life course perspective* (pp. 200–217). Cambridge University Press. https://doi.org/10.1017/9781108552684.013

Dutrévis, M., & Croizet, J.-C. (2005). Reputation of intellectual inferiority undermines memory efficiency among college students. *Current Research in Social Psychology, 10*, 104–115. https://psycnet.apa.org/record/2015-11200-001

Ferrand, L., & Alario, F. X. (1998). Normes d'associations verbales pour 366 noms d'objets concrets. *L'Année Psychologique, 98*(4), 659–709. https://doi.org/10.3406/PSY.1998.28564

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

Harkins, S. G. (2006). Mere effort as the mediator of the evaluation-performance relationship. *Journal of Personality and Social Psychology, 91*(3), 436–455. https://doi.org/10.1037/0022-3514.91.3.436

Hertzog, C., Touron, D. R., & Hines, J. C. (2007). Does a time-monitoring deficit influence older adults' delayed retrieval shift during skill acquisition? *Psychology and Aging, 22*(3), 607–624.

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

Hess, T. M., Emery, L., & Queen, T. L. (2009). Task demands moderate stereotype threat effects on memory performance. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 64*(4), 482–486.

Higgins, E. T. (1997). *Beyond pleasure and pain. American psychologist, 52*(12), 1280–1300.

Hummert, M. L. (2011). Age stereotypes and aging. In *Handbook of the psychology of aging* (7th ed., pp. 249–262). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-380882-0.00016-4

Jonker, C., Geerlings, M. I., & Schmand, B. (2000). Are memory complaints predictive for dementia? A review of clinical and population-based studies. *International Journal of Geriatric Psychiatry, 15*(11), 983–991. https://doi.org/10.1002/1099-1166(200011)15:11<983::AID-GPS238>3.0.CO;2-5

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

Maquestiaux, F., & Ruthruff, E. (2021). Testing the over-reliance on central attention (ORCA) hypothesis: Do older adults have difficulty automatizing especially easy tasks? *Journal of Experimental Psychology: General, 150*(9), 1722–1740. https://doi.org/10.1037/XGE0001020

Mazerolle, M., Régner, I., Morisset, P., Rigalleau, F., & Huguet, P. (2012). Stereotype threat strengthens automatic recall and undermines controlled processes in older adults. *Psychological Science, 23*(7), 723–727. https://doi.org/10.1177/0956797612437607

Mazerolle, M., Régner, I., Rigalleau, F., & Huguet, P. (2015). Stereotype threat alters the subjective experience of memory. *Experimental Psychology, 62*(6), 395–402. https://doi.org/10.1027/1618-3169/a000303

Mazerolle, M., Smith, A. M., Torrance, M., & Thomas, A. K. (2021). Understanding older adults' memory distortion in the light of stereotype threat. *Frontiers in Psychology, 12*, 628696. https://doi.org/10.3389/fpsyg.2021.628696

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning Memory and Cognition, 26*(5), 1170–1187. https://doi.org/10.1037/0278-7393.26.5.1170

O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin, 29*(6), 782–789. https://doi.org/10.1177/0146167203029006010

Popham, L. E., & Hess, T. M. (2015). Age differences in the underlying mechanisms of stereotype threat effects. *The Journals of Gerontology: Series B, 70*(2), 223–232. https://doi.org/10.1093/GERONB/GBT093

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

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

Seibt, B., & Förster, J. (2004). Stereotype threat and performance: How self-stereotypes influence processing by inducing regulatory foci. *Journal of Personality and Social Psychology, 87*(1), 38–56. https://doi.org/10.1037/0022-3514.87.1.38

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

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

Steele, C. M., & Davies, P. G. (2003). Stereotype threat and employment testing: A commentary. *Human Performance, 16*(3), 311–326. https://doi.org/10.1207/S15327043HUP1603_7

Touron, D. R. (2015). Memory avoidance by older adults. *Current Directions in Psychological Science, 24*(3), 170–176. https://doi.org/10.1177/0963721414563730

Touron, D. R., & Hertzog, C. (2004a). Distinguishing Age Differences in Knowledge, Strategy Use, and Confidence During Strategic Skill Acquisition. *Psychology and Aging, 19*(3), 452–466. https://doi.org/10.1037/0882-7974.19.3.452

Touron, D. R., & Hertzog, C. (2004b). Strategy shift affordance and strategy choice in young and older adults. *Memory & Cognition, 32*(2), 298–310.

Zacks, R. T., Hasher, L., & Li, K. Z. H. (2000). Human memory. In *The handbook of aging and cognition* (2nd ed., pp. 293–357). Lawrence Erlbaum Associates Publishers.

Zajonc, R. B. (1965). Social facilitation. *Science, 149*(3681), 269–274. https://doi.org/10.1126/SCIENCE.149.3681.269/ASSET/EF516A96-1362-4569-B7A3-6AE35EEAB24E/ASSETS/SCIENCE.149.3681.269.FP.PNG

Zhang, X., Fung, X. H. H., Stanley, X. J. T., Isaacowitz, X. D. M., & Ho, M. Y. (2013). Perspective taking in older age revisited: A motivational perspective. *Developmental Psychology, 49*(10), 1848–1858. https://doi.org/10.1037/A0031211