Psychology and Aging

Age-Based Stereotype Threat Mechanisms: Evidence From a Working-Memory Task

Sarah J. Barber, Caroline McDonald, Kehinde Akintola, and Dobromir Rahnev

Online First Publication, July 16, 2026. https://dx.doi.org/10.1037/pag0001006

CITATION

Barber, S. J., McDonald, C., Akintola, K., & Rahnev, D. (2026). Age-based stereotype threat mechanisms: Evidence from a working-memory task. *Psychology and Aging*. Advance online publication. https://dx.doi.org/10.1037/pag0001006

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychology and Aging

ISSN: 0882-7974

https://doi.org/10.1037/pag0001006

Age-Based Stereotype Threat Mechanisms: Evidence From a Working-Memory Task

Sarah J. Barber[1, 2], Caroline McDonald[1], Kehinde Akintola[1, 2], and Dobromir Rahnev[3]

[1] Department of Psychology, Georgia State University
[2] Gerontology Institute, Georgia State University
[3] School of Psychology, Georgia Institute of Technology

Age-based stereotype threat (ABST) occurs when older adults worry about being judged by a negative ageist stereotype. Although ABST can impair cognitive performance, its underlying mechanisms remain unclear, and two prominent frameworks offer different explanations. The *integrated process model* proposes that ABST heightens arousal, negative affect, and self-monitoring, which together consume working-memory resources and reduce processing efficiency. In a working-memory task, ABST should therefore lead to reduced hits, slower responding, increased false alarms, and slower evidence accumulation (i.e., lower drift rates in a drift diffusion model). In contrast, the *regulatory focus account* proposes that ABST induces a prevention focus that prioritizes vigilance and avoids commission errors. ABST should therefore lead to reduced hits, slower responding, reduced false alarms, increased omissions, and increased response caution (i.e., higher boundary separation in a drift diffusion model). To test these predictions, older adults completed a baseline 2-back task and then repeated the task under either control instructions or instructions designed to elicit ABST. Compared with the control condition, participants in the ABST condition showed reduced practice-related gains in hit rates and a selective increase in omissions. No ABST effects were observed for response times, false alarms, drift rates, or boundary separation. Although this pattern aligns with some aspects of the regulatory focus account, ABST did not consistently result in more cautious responding, and the ABST-related increase in omissions could also reflect reduced working-memory efficiency. Overall, the findings suggest that ABST can undermine correct responding without producing clear changes in underlying decision parameters.

Public Significance Statement

Exposure to negative age-related stereotypes can inadvertently impair older adults' performance on cognitive assessments. In this study, we found that when a memory test was framed as measuring age-related cognitive decline, older adults were less likely to improve with practice and were more likely to leave questions unanswered. These findings underscore the importance of creating testing environments that minimize stereotype-related pressure to ensure that older adults' cognitive abilities are evaluated accurately.

Keywords: stereotype threat, regulatory focus, drift diffusion model, older adults

Supplemental materials: https://doi.org/10.1037/pag0001006.supp

David Weiss served as action editor.

Sarah J. Barber https://orcid.org/0000-0002-4401-8902

The data reported here have not been previously published and are publicly available on Open Science Framework: https://osf.io/um8zd/overview?view_only=af5d826257784cd397d21b6419c0e5f6. The authors have no conflicts of interest to declare. Research reported in this publication was supported by the National Institute on Aging of the National Institutes of Health under Grant R03AG083314. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors thanks Supriya Mandaloju, Soraya Merriweather, Preia Montgomery, and Katie Valles for research assistance.

Generative artificial intelligence software (ChatGPT) was used only to refine wording and improve clarity of expression. It was not used to generate new content or conduct analyses. All artificial intelligence-suggested edits were carefully reviewed by the authors to ensure their accuracy.

Sarah J. Barber played a lead role in conceptualization, data curation, formal analysis, funding acquisition, methodology, resources, supervision, visualization, and writing–original draft, a supporting role in investigation, and an equal role in project administration. Caroline McDonald played a lead role in investigation, a supporting role in data curation and writing–review and editing, and an equal role in project administration. Kehinde Akintola played a supporting role in data curation and project administration and an equal role in investigation. Dobromir Rahnev played a supporting role in conceptualization, formal analysis, funding acquisition, and methodology and an equal role in writing–review and editing.

Correspondence concerning this article should be addressed to Sarah J. Barber, Department of Psychology, Georgia State University, P.O. Box 5010, Atlanta, GA 30302, United States. Email: sbarber10@gsu.edu

Older adults typically perform worse than younger adults on tests of cognitive abilities (e.g., Old & Naveh-Benjamin, 2008; Salthouse, 2004; Verhaeghen et al., 1993). However, negative expectations about aging can exacerbate these deficits via stereotype threat. Stereotype threat is a disruptive concern that occurs when people know that poor performance on their part will confirm a negative self-relevant stereotype. When experiencing stereotype threat, people often underperform compared with their potential (Steele & Aronson, 1995; for a review, see Spencer et al., 2016). For instance, research shows that when negative stereotypes about age-related cognitive decline are made salient, older adults underperform on memory tests (Hess et al., 2003; for a review, see Barber & Lui, 2020) and these effects have important clinical health implications (for reviews, see Barber, 2020; Régner & Huguet, 2025). For instance, age-based stereotype threat (ABST)[1] can impair older adults' performance on brief cognitive tests used in clinical settings to evaluate for cognitive impairment (Barber et al., 2015; Fresson et al., 2017; Haslam et al., 2012) and can increase the false positive rate on these tests (Mazerolle et al., 2017). The overarching goal of the present study was to better understand the mechanisms through which ABST leads to these cognitive performance impairments.

Stereotype Threat: Theoretical Frameworks

Despite extensive research on the mechanisms underlying stereotype threat effects, the precise processes responsible for stereotype threat-related performance deficits remain contested. One influential account is the integrated process model (Schmader et al., 2008), which proposes that stereotype threat disrupts performance through several interrelated mechanisms. First, stereotype threat is proposed to increase sympathetic arousal, which results in a direct impairment of prefrontal processing and hence results in poorer cognitive performance. In addition, stereotype threat is proposed to create negative mood states that people try to suppress while also increasing performance monitoring. Because these processes are cognitively demanding, stereotype threat reduces the amount of working-memory resources available to perform the critical task, and hence performance suffers. Thus, in this model, reduced working-memory efficiency serves as the common distal mediator that accounts for stereotype threat-related performance deficits.

Consistent with the integrated process model, several studies with older adult participants have shown that ABST effects are mediated by increases in anxiety (Abrams et al., 2006, 2008; Swift et al., 2013). Additional support for this model comes from research showing that ABST effects are more likely to emerge on cognitive tests that are more heavily reliant on working-memory resources. For instance, ABST affects controlled rather than automatic processes of cognition (Mazerolle et al., 2012), exerts a greater detrimental impact on working memory than on episodic memory (Armstrong et al., 2017), and more strongly affects associative memory as compared with item memory performance (Brubaker & Naveh-Benjamin, 2018).

While the integrated process model emphasizes working-memory interference as the core mechanism underlying stereotype threat effects, other perspectives focus primarily on motivational factors. One such perspective draws on regulatory focus theory, which is a motivational framework that distinguishes between promotion and prevention orientations (Higgins, 1997, 1998). Individuals with a promotion orientation are concerned with accomplishment and improvement and pursue their goals eagerly by approaching gains (i.e., striving to make good things happen). In contrast, individuals with a prevention orientation are focused on responsibilities and safety, fulfilling their duties vigilantly by avoiding risks and losses (i.e., striving to prevent bad things from happening). Although people differ in their chronic tendency to be promotion or prevention oriented, a given individual's regulatory focus can also vary across situations (Grimm et al., 2012), and stereotype threat is proposed to be a situational factor that can shift individuals toward a prevention-oriented motivational state (Seibt & Förster, 2004; see also Barber, 2017; Barber & Mather, 2013a; Gaillard et al., 2011; Popham & Hess, 2015). More specifically, when individuals encounter a negative self-relevant stereotype, it sets up a negative reference point in which the best possible outcome is avoidance of failure rather than the achievement of success. This negative reference point is proposed to increase vigilance to avoid losses and errors of commission, leading individuals to adopt a short-term prevention orientation (even when their dispositional tendency is a promotion orientation). Importantly, while this account predicts that a prevention focus induced by stereotype threat may alter behavior, it does not necessarily assume that this is due to reductions in working-memory resources.

If stereotype threat induces a short-term prevention focus, then older adults' performance in threat-eliciting situations should be characterized by slower and more cautious responding characterized by fewer errors of commission. Consistent with this hypothesis, research has shown that ABST can induce a more conservative response criteria and reduce older adults' memory false alarms (Barber & Mather, 2013b; Rossi-Arnaud et al., 2018; Wong & Gallo, 2016). ABST also leads older adults to respond more slowly (Abrams et al., 2006; Popham & Hess, 2015; Zhang et al., 2017), to withhold low-confidence responses (Rossi-Arnaud et al., 2018; Thomas et al., 2020), and to be more risk averse in decision making (Coudin & Alexopoulos, 2010). Recent neuroimaging research also demonstrates that ABST is associated with increased activity in parietal midline regions (such as the posterior cingulate cortex and precuneus) that are associated with an error-prevention focus (Chen et al., 2022).

In summary, within the existing literature, there is empirical support for both the integrated process model and the regulatory focus model of ABST effects among older adults. However, neither framework fully accounts for the range of findings reported to date. To advance our understanding of the mechanisms underlying ABST effects, the present study examines how ABST impacts the component processes of working-memory task performance (i.e., hits, false alarms, omissions, and response times [RTs]) and uses a well-validated computational approach (i.e., the drift diffusion model [DDM]) to further decompose ABST effects.

The DDM

According to Ratcliff's diffusion model for two-alternative forced choice tasks (Ratcliff, 1978), when people are presented with an item and are asked to make a speeded two-choice decision (e.g., Is this item old or new?), they accumulate evidence continuously

[1] While the term *stereotype threat* is used to describe performance decrements that occur when individuals fear confirming a negative stereotype about a social group to which they belong, the term *ABST* refers more specifically to stereotype threat arising from concerns about confirming negative stereotypes related to age.

toward one of two response boundaries. Ultimately, a response of "old" is made if the accumulated evidence reaches the upper boundary, whereas a response of "new" is made if the accumulated evidence reaches the lower boundary. Importantly, both the rate at which evidence accumulates and the distance between decision boundaries can vary across individuals and situations, reflecting differences in processing efficiency and response caution. DDM allows these latent parameters to be estimated quantitatively (Ratcliff et al., 2016; Voss et al., 2004). Specifically, the boundary separation parameter (a) reflects the amount of evidence required before making a decision. Smaller separations between boundaries indicate a preference for speed over accuracy, whereas larger separations indicate a more cautious or conservative decision strategy. The drift rate (v) represents the average quality of evidence accumulation, with higher drift rates indicating more efficient information processing and yielding faster, more accurate responses. Although drift rate is conceptually related to the signal detection parameter d', it also incorporates decision speed, providing a more dynamic measure of sensitivity per unit of time (Busemeyer & Diederich, 2010).

DDM has been widely applied to study age-related changes in cognitive performance (Ging-Jehli & Ratcliff, 2020; Ratcliff et al., 2004, 2007; Spaniol et al., 2006; Theisen et al., 2021). However, no prior research has applied this model to understand ABST, making the present study the first to do so. Because DDM estimates latent decision processes from the joint distribution of RTs and accuracy, tasks used for this approach typically require many speeded decisions to obtain reliable parameter estimates. Accordingly, the present study examined the effects of ABST on performance in a working-memory task that included a 1,500-ms response deadline, which was intended to encourage rapid responding and support reliable estimation of the DDM parameters.

Using DDM provides several advantages for examining ABST mechanisms. For example, prior work has shown that ABST is often associated with slower RTs, but it remains unclear whether this reflects poorer information encoding (as proposed by the integrated process model) or greater response caution (as proposed by the regulatory focus account). DDM can disentangle these alternatives by estimating whether changes in performance are driven by drift rate (information quality) or boundary separation (decision threshold).

Overview of Present Study

Although the diffusion model was originally formulated for recognition memory (Ratcliff, 1978), it has since been applied to a range of binary decision paradigms, including working-memory tasks with young adults (Pedersen et al., 2023; Sewell et al., 2016; Shepherdson et al., 2018) and older adult participants (Jang et al., 2022). The present study focuses on how ABST impacts older adults' working-memory performance, given evidence that ABST effects are most robust in this domain (Armstrong et al., 2017).

According to the integrated process model, ABST increases physiological arousal and task monitoring demands while also inducing negative mood states that individuals attempt to regulate. Together, these processes consume limited working-memory resources (Schmader et al., 2008), which should manifest as lower overall accuracy, fewer hits, more false alarms, and slower responses during a working-memory task. With fewer cognitive resources available for task performance, memory representations should become less precise, resulting in a reduced drift rate (v).

In contrast, the regulatory focus model proposes that ABST induces a prevention focus, heightening vigilance and encouraging a cautious response strategy aimed at avoiding mistakes (Seibt & Förster, 2004). This motivational shift should lead to slower RTs, with fewer hits but also with fewer false alarms. ABST should also lead to more errors of omission, as individuals become more conservative and are more likely to withhold low-confidence responses. Finally, this account also suggests that ABST should be associated with a larger boundary separation (a), reflecting a stronger preference for accuracy over speed.

Method

Design and Participants

This experiment used a 2 (Instructional Condition: ABST vs. control) × 2 (Task Phase: baseline vs. critical) mixed design, with instructional condition manipulated between subjects and task phase manipulated within subjects. Sample size was determined using a power analysis conducted with G*Power 3.1.9.7. Within this analysis, we assumed a medium effect size of d = .373 (equivalent to f = .186), which is the meta-analytic average effect size of ABST on working memory (Armstrong et al., 2017). We also specified an α = .05 and assumed a correlation of 0.50 between our repeated measures. Based upon this, we determined that a sample of 78 participants would provide over 90% power to detect the expected within-between interaction. To account for potential data exclusions, we aimed to recruit a larger sample than this target.

Participants were recruited through multiple channels, including flyers distributed at local senior centers and community organizations, outreach to individuals in an existing database of adults who had previously expressed interest in aging-related research, targeted social media advertisements, and email invitations sent to current Georgia State University students and alumni aged 60 years and older. We also advertised this study on ResearchMatch, which is a national registry funded by the U.S. National Institutes of Health that facilitates participant recruitment for research studies.

Interested individuals completed a brief screening to determine eligibility. Only English-speaking individuals between the ages of 60 and 85 years were eligible to participate. To ensure that participants were free from conditions that could affect cognitive performance, exclusion criteria also included (a) self-reported history of neurological conditions; (b) an indication of clinically relevant depression symptoms, defined as a score of 6 or more on the Geriatric Depression Scale Short Form (Brink et al., 1982); (c) hospitalization within the prior 6 months; and (d) a self-reported diagnosis of cognitive impairment. Participants were also excluded if they had uncorrected visual impairments that would interfere with reading information on a computer screen or were unable to make unassisted keyboard responses.

In total, 165 individuals were screened for eligibility. Of these, 22 did not meet the inclusion criteria, and an additional three declined to participate after completing the screening process. Of the remaining 140 eligible individuals, 117 scheduled a testing session. The testing sessions took place between September 2024 and June 2025. Of the 117 individuals who completed the study, based upon our preregistered data analysis plan, we subsequently excluded eight participants with below-chance performance on the baseline working-memory task. Although not included in our preregistration, we also opted to exclude an additional five participants: three participants for failing to provide a response to 25% or more of the baseline task trials and two for not understanding the task instructions. This left a final sample of 93 participants (M_{age} = 68.26; range = 60–82 years), including 27 men and 66 women. The final sample was racially diverse, with 54.8% identifying as White, 38.7% identifying as Black, 5.5% identifying as biracial or multiracial, and one individual declining to specify their race. Within the final sample, there were 48 participants in the control condition and 45 participants in the ABST condition. There were no significant differences between participants in the control and ABST groups in any of the assessed demographic or self-rated health variables (see Table 1).

Table 1

Mean Scores (and Standard Deviations) for Demographic and Health Measures as a Function of Task Instruction Condition

Demographic and health measure	Control condition (n = 48)	ABST condition (n = 45)
Age (years)	66.64 (5.21)	67.62 (5.26)
Gender		
Male	14	13
Female	34	32
Race		
Caucasian/White	27	24
African American/Black	18	18
Biracial/multiracial	2	3
Declined to state	1	0
Education (years)	16.32 (2.32)	16.75 (2.64)
Employment status		
Employed full time	7	4
Employed part time	4	8
Unemployed[a]	8	9
Retired	28	23
Declined to state	1	1
Subjective social status[b]	6.46 (1.35)	6.70 (1.36)
Subjective cognitive difficulties[c]		
Memory	4.57 (0.83)	4.59 (0.90)
Executive function	4.74 (0.94)	4.98 (0.88)
Attention	4.64 (0.92)	4.89 (0.87)
SF-36 health survey scores[d]		
Physical functioning	80.11 (21.30)	82.50 (20.98)
Bodily pain	75.16 (21.27)	76.31 (22.14)
Role limitations due to health	80.85 (33.06)	80.68 (33.61)
Role limitations due to emotional problems	86.52 (30.82)	89.39 (23.59)
Emotional well-being	81.79 (12.78)	82.68 (11.60)
Social functioning	85.90 (22.81)	92.61 (15.79)
Energy/fatigue	66.38 (15.59)	68.98 (16.86)
General health perceptions	69.68 (18.16)	72.39 (16.62)
Dispositional regulatory focus[e]		
Promotion focus	3.89 (1.48)	3.83 (1.18)
Prevention focus	6.80 (1.44)	6.80 (1.61)

Note. All data reported here come from the pre-experimental survey. One participant in the control condition and one participant in the ABST condition did not complete this survey; for these individuals, only basic demographic information (age, gender, and race) is available. ABST = age-based stereotype threat.
[a] The category of "unemployed" included individuals who reported being unemployed and looking for work, unemployed and not looking for work, unable to work, or enrolled as full-time students. [b] Subjective social status was assessed using the McArthur Scale of Subjective Social Status (Adler et al., 2000). Participants were shown a drawing of a ladder with 10 rungs and were told the ladder represents where people stand in society. Participants indicated where they themselves stand on the ladder, with higher scores indicating higher subjective social status. [c] Subjective cognitive difficulties were assessed using a three-item questionnaire from Opdebeeck et al. (2019). Participants rated their day-to-day memory abilities, executive functioning, and attention relative to others their age on a 1 (*very poor*) to 6 (*excellent scale*) Likert scale. [d] Health status was assessed using the RAND 36-Item Short Form Health Survey (SF-36), which provides measures of eight health concepts. Each of the eight domain scores are transformed to a 0–100 scale, with higher scores representing better health within that domain (Hays et al., 1993). [e] Dispositional regulatory focus was assessed using a modified version of the General Regulatory Focus Measure (Lockwood et al., 2002). This 18-item questionnaire includes questions assessing both a prevention orientation (e.g., *In general I am focused on preventing negative events in my life*) and a promotion orientation (e.g., *I frequently imagine how I will achieve my hopes and aspirations*). Questions are answered using a 1 (*not at all true of me*) to 9 (*very true of me*) Likert scale. The word "academic" was removed from two items to make the questions more accessible to older adults. This measure was included as a baseline descriptive variable to confirm that participants in the ABST and control conditions did not differ in dispositional regulatory focus prior to the experimental manipulation. However, because the regulatory focus account posits that ABST induces a transient prevention focus regardless of an individual's dispositional orientation, these scores were not included as moderators in the analyses.

Procedure

The study consisted of two parts: an online pre-experimental survey followed by an in-person laboratory session.

Pre-Experimental Survey

Following eligibility screening and prior to the laboratory session, participants first completed the online survey at the time and location of their choosing (although two participants, one from each condition, failed to complete this survey). At its outset, participants provided informed consent before completing a series of questionnaires, including those assessing demographic characteristics and health status (see Table 1). A full list of all measures included in the pre-experimental survey is provided in the Supplemental Materials. There were no significant differences between participants in the two conditions on any measure included in the pre-experimental survey.

Experimental Session

The in-person experimental session took place either at Georgia State University or at a local senior center in Atlanta, Georgia. At the start of each session, participants provided informed consent. Participants were then assigned to one of the two between-subject conditions using stratified randomization, ensuring that each condition included approximately equal numbers of men and women and White and Black participants.

We then proceeded to assess working-memory performance using n-back tasks. To familiarize participants with the task, they first completed a practice block of a 1-back task. During this practice task, participants viewed a series of digits (1–9) presented sequentially in white font on a black background in the center of the computer screen. Each digit appeared for 1,500 ms, followed by feedback displayed for 500 ms and an interstimulus interval of 750 ms.[2] Of note, the response window was limited to 1,500 ms to encourage speeded responding, which is necessary for estimating the DDM parameters from RT distributions. On each trial, participants indicated whether the current digit was a "match" or "nonmatch" relative to the digit presented on the immediately preceding trial. When the digit was a match, participants pressed the S key ("same" as the preceding number); when it was a nonmatch, they pressed the N key ("not" the same). Trials in which participants responded incorrectly or failed to respond within the 1,500-ms limit were considered incorrect and were followed by a buzzer sound during the feedback period, whereas correct responses were followed by a chime sound. The practice block consisted of 22 trials, with feedback provided after each response.

After completing the practice 1-back task, participants next completed a practice 2-back task. This was methodologically identical to the practice 1-back task except for the level of difficulty. In this version of the task, participants indicated whether the current digit matched the one presented two trials earlier. The practice 2-back also included 22 trials, with feedback provided after each response.

After completing two practice tasks, participants performed a 2-back task that served as the baseline measure of working-memory performance. This task was methodologically identical to the practice 2-back task, except that no feedback was provided. The baseline 2-back task consisted of 152 trials, with the first two trials excluded from analyses. Of the remaining 150 trials, 75 were "match" trials (the current digit matched the one presented two trials earlier), and 75 were "nonmatch" trials (the current digit did not match the one presented two trials earlier). After completing the baseline working-memory task, participants were given a 5-min break, and during this time, they were left alone in the testing room.

Participants then completed the 2-back task a second time, which served as the critical working-memory task. This task was methodologically identical to the baseline 2-back task, but the instructions provided at the outset of this task varied depending on participant's randomly assigned task condition. Participants in the control condition were simply informed that they would be completing the 2-back task again. In contrast, participants in the ABST condition received instructions designed to make negative age-related stereotypes about memory performance salient. Consistent with common experimental approaches used in ABST research, these instructions included several elements intended to activate ABST, including framing the task as diagnostic of age-related memory decline, identifying participants as members of the older adult group, and referencing a comparison with younger adults. Specifically, participants in the ABST condition were presented with the following instructions on the computer screen, which were also read aloud by the experimenter:

> Before you begin, we would like to explain more about today's experiment. *The purpose of this next test is to investigate age-related declines in memory performance* [emphasis added]. Based upon your age, you are part of our *older adult group*. Your performance will be compared to that of younger adults. The current research will help us understand why older adults—like yourself—experience memory declines compared to younger adults.

After receiving the instructions associated with their randomly assigned condition, participants proceeded to complete the critical 2-back task. This was followed by a series of additional questionnaires and cognitive tasks, which will not be further discussed here. Finally, at the end of the session, the experimenter explained the purpose of the study and answered participants' questions. As part of this debriefing, the experimenter disclosed our use of mild deception to the participants in the ABST condition. Specifically, these participants had been told that their performance would be compared with that of younger adult participants, but this was not true. Participants were informed that this instruction was instead designed to induce a feeling of ABST and that the purpose of the study was to evaluate how this impacted their performance.

Measures

Overall accuracy was calculated as the proportion of correct responses across all 150 trials, and this was done separately for the baseline and critical 2-back tasks. To provide a more detailed characterization of performance, task accuracy was further decomposed into hit rates, false alarm rates, and omissions. A hit was defined as a correct response on a "match" trial (i.e., correctly identifying a digit as the same as the one presented two trials earlier), whereas a false alarm was defined as an incorrect identification of a

[2] The presentation timing during this task was based upon pilot testing conducted with nine additional participants.

"nonmatch" trial as a match. Omissions referred to trials in which participants failed to respond within the 1,500-ms response window. In addition to accuracy-based measures, RT was calculated for both the baseline and critical tasks. RTs shorter than 200 ms were excluded from these calculations.

We also estimated the DDM parameters drift rate (v) and boundary separation (a) using the EZ-diffusion method (Wagenmakers et al., 2007), which is a robust approach for estimating individual differences in these parameters (van Ravenzwaaij & Oberauer, 2009). This approach derives parameter estimates for each participant using three summary statistics from their task performance: overall mean accuracy, mean RT for correct trials, and the variance of RTs for correct trials.[3]

Analytic Plan

All analyses were conducted using IBM SPSS Statistics (Version 29), with an α level of .05 for all significance tests. To test our hypotheses, we conducted a series of 2 (Instructional Condition: ABST vs. control) × 2 (Task Phase: baseline vs. critical) analyses of variance (ANOVAs) for each dependent measure. The primary effect of interest for each analysis was the Instructional Condition × Task Phase interaction. When significant interactions were observed, simple-effects tests were conducted within each condition to clarify the direction and magnitude of the effects.

Statistical assumptions for the ANOVA analyses were met for each outcome measure, except for the omissions. Because omissions were non-normally distributed and positively skewed, they violated the assumptions of ANOVA. Although this deviated from our preregistration, we therefore analyzed omission counts using a generalized linear mixed model with a Poisson distribution and log link function, which is well suited for count data. The model included instructional condition (ABST vs. control) as a between-subjects factor and task phase (baseline vs. critical) as a within-subjects factor, with a random intercept for participant to account for repeated observations. This approach accommodated individual differences in baseline omission rates and tested how omission counts varied across conditions and task phases. When a significant Instructional Condition × Task Phase interaction emerged, pairwise comparisons of the estimated marginal means were conducted to clarify the direction and magnitude of the effect while accounting for both fixed and random effects in the model.

Within these analyses, a hypothesis that ABST would *increase* a parameter (e.g., boundary separation) would be supported if the ABST group showed a larger increase (or a smaller decrease) from the baseline to the critical task relative to the control group. Conversely, a hypothesis that ABST would *decrease* a parameter (e.g., drift rate) would be supported if the ABST group showed a larger decrease (or a smaller increase) from the baseline to the critical task relative to the control group. For example, both theoretical models predict that ABST should decrease hit rates. This could manifest as a decline in hit rates from the baseline to the critical task in the ABST condition but not in the control condition. However, it is also possible that practice effects could occur, such that participants (regardless of instructional condition) exhibit an increase in hit rates from the baseline to the critical task. In this case, we would expect a smaller increase in hit rates for participants in the ABST condition relative to the control condition. In other words, although both groups might show practice-related gains, these gains would be attenuated for those in the ABST group, producing the predicted Instructional Condition × Task Phase interaction.

Transparency and Openness

All study procedures were approved by the Georgia State University Institutional Review Board (Protocol No. H23668). A preregistration of the study hypotheses, procedures, and analysis plan was created on the Open Science Framework (https://osf.io/um8zd/overview?view_only=af5d826257784cd397d21b6419c0e5f6) in August 2024, prior to data collection. However, due to experimenter error, the preregistration was inadvertently saved as a draft and was not formally registered. This error was not discovered until October 2025, after data collection had been completed and preliminary analyses had begun. At that time, minor clarifications were added to the preregistration draft (e.g., details about exploratory analyses and the method for calculating diffusion model parameters), but the planned sample size, primary hypotheses, and data exclusion criteria remained unchanged from those specified in the August 2024 draft. Unless otherwise noted, all analyses reported here followed the preregistered plan. The data analyzed in this article have not been published previously. Deidentified data are publicly available at https://osf.io/um8zd/overview?view_only=af5d826257784cd397d21b6419c0e5f6. The experimental program is available upon request from the corresponding author.

Results

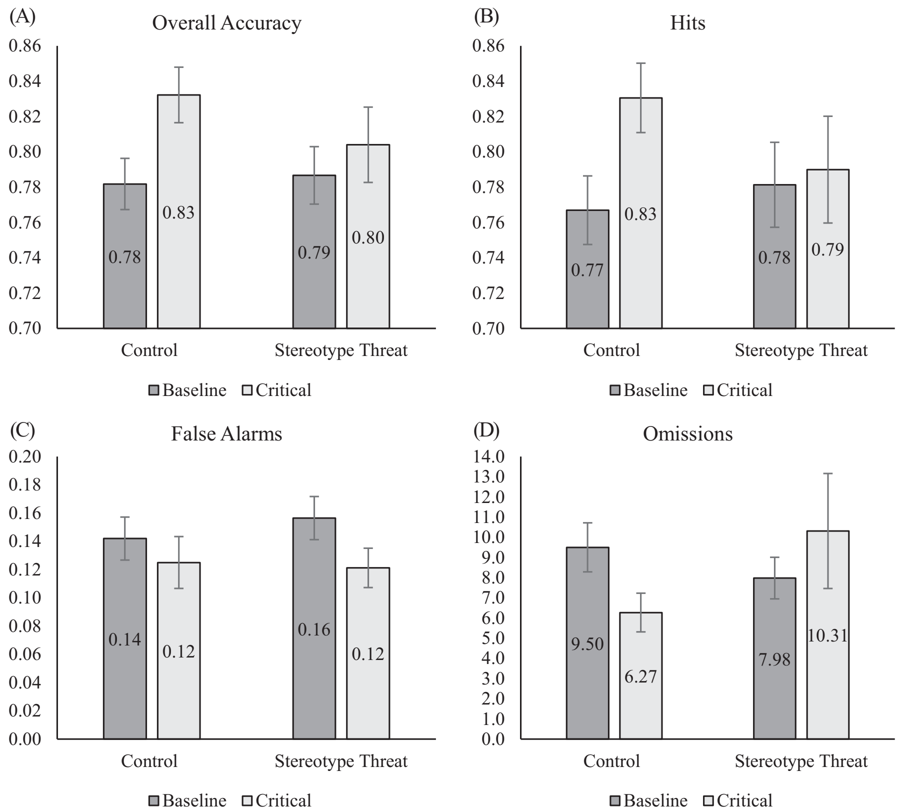
Accuracy-Based Measures

We first examined whether changes in overall accuracy from the baseline to the critical task differed as a function of task condition. A 2 (Instructional Condition: ABST vs. control) × 2 (Task Phase: baseline vs. critical) mixed-factor ANOVA revealed a significant main effect of task phase, $F(1, 91) = 17.19$, $p < .001$, $\eta_p^2 = .159$, with accuracy increasing from the baseline to the critical task. Although there was no significant main effect of instructional condition, $F(1, 91) = 0.27$, $p = .608$, $\eta_p^2 = .003$, the interaction between instructional condition and task phase was statistically significant, $F(1, 91) = 4.10$, $p = .046$, $\eta_p^2 = .043$. As shown in Figure 1A, follow-up comparisons indicated that, for participants in the control group, overall accuracy increased significantly from the baseline ($M = .78$) to the critical task ($M = .83$), $t(47) = 5.78$, $p < .001$, $d = -.834$. In contrast, participants in the ABST showed no significant change in overall accuracy between the baseline ($M = .79$) and critical tasks ($M = .80$), $t(44) = 1.23$, $p = .225$, $d = -.184$. Thus, practice-related improvements in accuracy were attenuated in the ABST condition. This pattern mirrors prior research demonstrating that ABST adversely affects working-memory performance in older adults.

[3] The EZ-diffusion method (Wagenmakers et al., 2007) does not provide explicit guidance for handling omissions. However, because nonresponses lack both a choice and a RT, they cannot easily be incorporated into the diffusion model estimates. Consistent with prior recommendations (Boag et al., 2025), omissions were therefore excluded from the EZ-diffusion calculations (i.e., accuracy for EZ diffusion calculations = correct responses/[correct responses + incorrect responses]). In contrast, when computing overall accuracy for the behavioral analyses reported in the Results section, omissions were included in the denominator (i.e., overall accuracy = correct responses/total number of trials).

Figure 1

Working-Memory Task Accuracy as a Function of Task Phase and Instructional Condition



Note. Overall accuracy (A), hit rate (B), and false alarms (C) are presented as proportions. Omissions (D) are presented as counts, as their distribution was non-normal and positively skewed; therefore, analyses were conducted on omission counts rather than proportions. Errors bars represent the standard errors of the mean.

To identify the specific aspects of performance that contributed to the reduced improvement of overall accuracy in the ABST instruction condition, we conducted separate 2 (Instructional Condition: ABST vs. control) × 2 (Task Phase: baseline vs. critical) mixed-factor ANOVAs for hits, false alarms, and omissions. In the analysis focused on hit rates, there was a significant main effect of task phase, $F(1, 91) = 9.63$, $p = .003$, $\eta_p^2 = .096$, which was due to hit rates increasing from the baseline to critical task. While there was no significant main effect of instructional condition, $F(1, 91) = 0.18$, $p = .675$, $\eta_p^2 = .002$, the interaction between instructional condition and task phase was statistically significant, $F(1, 91) = 5.59$, $p = .020$, $\eta_p^2 = .058$. As shown in Figure 1B, follow-up comparisons indicated that participants in the control group showed a significant increase in hit rate from the baseline ($M = .77$) to the critical task ($M = .83$), $t(47) = 5.18$, $p < .001$, $d = -.748$, whereas participants in the ABST group did not show a significant change ($M = .78$ vs. $M = .79$), $t(44) = 0.43$, $p = .672$, $d = -.064$.

Analysis of the false alarms revealed only a significant main effect of task phase, $F(1, 91) = 12.63$, $p < .001$, $\eta_p^2 = .096$, with false alarms decreasing from the baseline to critical task (see Figure 1C). However, there was no significant main effect of instructional condition, $F(1, 91) = 0.06$, $p = .802$, $\eta_p^2 = .001$, and no significant interaction between instructional condition and task phase, $F(1, 91) = 1.55$, $p = .216$, $\eta_p^2 = .017$, suggesting that the frequency of commission errors did not significantly differ between the ABST and control groups.

Omissions were analyzed using a generalized linear mixed model with a Poisson distribution and log link function. The model successfully converged, and the inclusion of a random intercept for participant substantially improved model fit, indicating strong clustering by participant (intraclass correlation coefficient = .86). Results revealed neither a significant main effect of instructional condition, $F(1, 182) = 0.003$, $p = .957$, nor a significant main effect of task phase, $F(1, 182) = 2.41$, $p = .122$. However, as shown in Figure 1D, there was a significant interaction between instructional condition and task phase, $F(1, 182) = 43.18$, $p < .001$. Follow-up pairwise comparisons of the model-estimated means showed that within the control condition, omissions significantly *decreased* from the baseline to the critical task, $t(182) = -4.52$, $p < .001$. In contrast, within the ABST condition, omissions significantly *increased* from the baseline to the critical task, $t(182) = 3.27$, $p = .001$.

Response Time

For overall RTs, the 2 (Instructional Condition) × 2 (Task Phase) mixed-factor ANOVA revealed a significant main effect of task phase, $F(1, 91) = 28.39$, $p < .001$, $\eta_p^2 = .238$, which reflected the fact that participants in both groups responded faster during the critical task than during the baseline task. However, as shown in Figure 2A, there was neither a significant main effect of instructional condition, $F(1, 91) = 0.61$, $p = .438$, $\eta_p^2 = .007$, nor a significant interaction between instructional condition and task phase, $F(1, 91) = 1.51$, $p = .222$, $\eta_p^2 = .016$. The same pattern of results was observed when analyses were restricted to RTs for correct responses only.

DDM Parameters

Finally, we examined whether our stereotype threat manipulation influenced the underlying decision processes contributing to task performance, as indexed by the DDM parameters of drift rate (v) and boundary separation (a). To do so, separate 2 (Instructional Condition) × 2 (Task Phase) mixed-factor ANOVAs were conducted on these two parameters. Results revealed significant main effects of task phase for the analyses of both drift rate, $F(1, 91) = 44.25$, $p < .001$, $\eta_p^2 = .327$, and of boundary separation, $F(1, 91) = 18.10$, $p < .001$, $\eta_p^2 = .166$. As shown in Figure 2B and 2C, both parameters increased in magnitude from the baseline to the critical task. However, neither the main effects of instructional condition nor the interactions between instructional condition and task phase were significant in these analyses (all ps > .325).

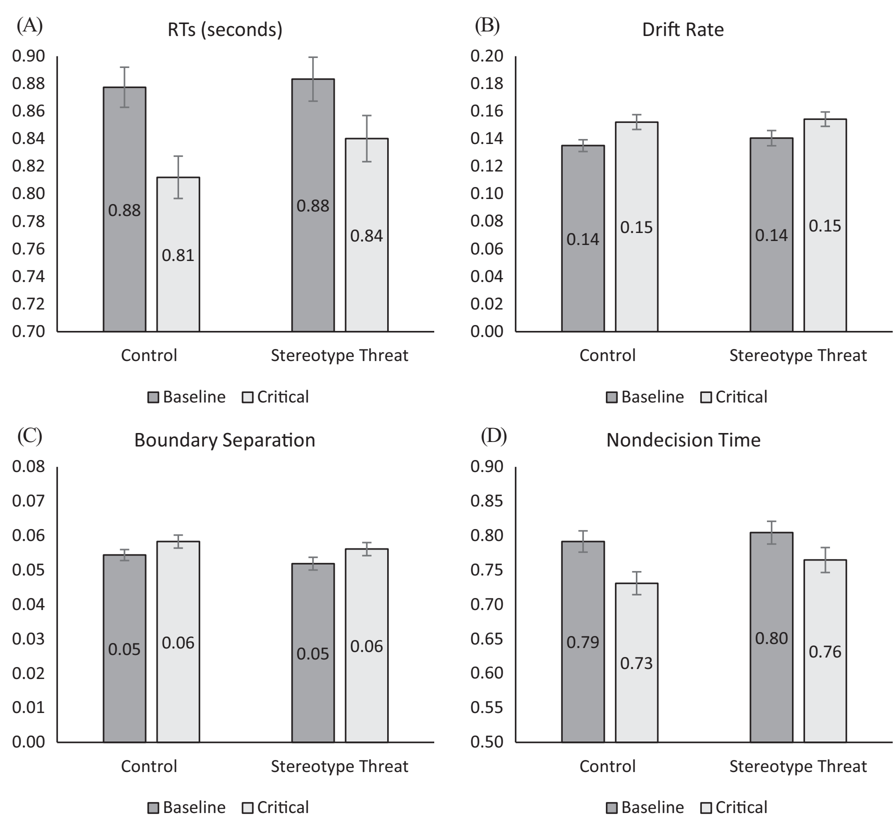
The EZ-diffusion method also provides an estimate of nondecision times (T_{er}). Although we did not preregister predictions about how ABST would influence this parameter, we conducted an exploratory 2 (Instructional Condition) × 2 (Task Phase) ANOVA on nondecision times. As shown in Figure 2D, this analysis revealed a significant main effect of task phase, $F(1, 91) = 35.14$, $p < .001$, $\eta_p^2 = .279$, indicating that nondecision time decreased from the baseline to the critical task. In contrast, neither the main effect of instructional condition nor its interaction with task phase was significant (both ps > .220).

Discussion

People are living longer than ever before, yet ageism and widely held negative stereotypes about aging continue to pose barriers to healthy cognitive and psychological functioning in later life (Barber et al., 2024). One specific challenge comes in the form of ABST, which occurs when older adults worry about the possibility of confirming (or being judged in light of) a negative ageist stereotype. When experiencing ABST, older adults often underperform on cognitive tasks, including brief screening measures commonly used in clinical contexts to detect cognitive decline or dementia (for reviews, see Barber, 2020; Barber & Lui, 2020). The present study

Figure 2

Working-Memory Task RTs and Drift Diffusion Model Parameters as a Function of Task Phase and Instructional Condition



Note. RTs (in seconds) are presented for all responses (A). The drift rate (B), boundary separation (C), and nondecision time (D) parameters were calculated using the EZ diffusion method. Errors bars represent the standard errors of the mean. RT = response time.

was designed to investigate the mechanisms underlying these ABST-related performance impairments. To do so, we randomly assigned older adult participants to either a control condition or to an ABST condition and assessed their performance on a working-memory task. In addition to examining overall accuracy, hit rates, false alarms, omissions, and RTs, we also employed a DDM to decompose performance into latent parameters reflecting the efficiency of evidence accumulation (i.e., the drift rate parameter) and the degree of response caution (i.e., the boundary separation parameter).

We tested competing theoretical accounts about how ABST should impact each of these aspects of older adults' performance. According to the integrated process model (Schmader et al., 2008), ABST impairs performance by consuming working-memory resources through increased physiological arousal, negative affect, and self-monitoring, thereby reducing the efficiency with which information is encoded and maintained. Reduced processing efficiency should, in turn, lead to less precise memory representations, manifesting as lower hit rates, greater false alarms, and slower responses. Within the DDM framework, the integrated process model further predicts that ABST should reduce drift rate, reflecting slower and less efficient evidence accumulation. In contrast, the regulatory focus account (Barber, 2017; Seibt & Förster, 2004) proposes that stereotype threat induces a motivational shift from a promotion to a prevention orientation, heightening vigilance and prioritizing accuracy over speed. This prevention focus should lead to a more conservative response strategy, reflected in slower RTs, fewer hits, fewer false alarms, and more omissions. Within the DDM framework, this account further predicts that ABST should be associated with an increased boundary separation parameter, indicating a more cautious decision threshold.

There are some common predictions made by both the integrated process model and by the regulatory focus account. Both frameworks predict that ABST should be associated with reduced overall accuracy, reduced hits, and slower responding. Only some of these shared predictions were supported. Participants in the ABST condition showed reduced improvement in overall accuracy and hit rates from the baseline to the critical task, consistent with the notion that threat undermines performance. However, ABST was not associated with slower RTs. While this pattern of results was not expected, it is important to note that all prior studies that have found an association between ABST and slower responding have used tasks that allowed for self-paced responding (Abrams et al., 2006; Popham & Hess, 2015; Zhang et al., 2017). In contrast, the present study imposed a 1,500-ms response deadline during the working-memory tasks. While imposing time constraints can amplify the detrimental effects of ABST on memory accuracy (Hess et al., 2009), it is also possible that the time constraints reduced variability in RTs, thereby constraining our ability to detect differences in RTs between the control and ABST conditions. Alternatively, participants in the ABST condition may have prioritized meeting the response deadline over accuracy, leading to maintained RTs but poorer performance.

While there were common predictions made by both the regulatory focus and integrated process model, the two theoretical accounts also diverge in some predictions. However, when comparing the differential predictions of the two accounts, the findings from this study do not provide clear support for one framework over the other. Looking first at the integrated process model, there was no evidence in the present study supporting its unique predictions. This model posits that stereotype threat reduces working-memory resources. As a result, ABST should be associated with a higher false alarm rate, due to reduced precision of memory representations, but this was not observed. This framework also predicts that ABST should be characterized by slower evidence accumulation, reflected in lower drift rates. However, this effect was also not observed.

At the same time, there was no consistent evidence in favor of the regulatory focus account's unique predictions. On the one hand, the pattern of omissions aligned with predictions based upon the regulatory focus account. For participants in the ABST condition, omissions significantly increased from the baseline to the critical task, whereas participants in the control condition showed a significant decrease. This selective increase in omissions under ABST is consistent with the idea that threat can promote a more cautious, prevention-oriented response style that emphasizes avoiding mistakes over maximizing hits. On the other hand, other predictions based upon the regulatory focus account were not supported. If ABST induces more cautious responding, it should also have led to fewer false alarms and a corresponding increase in the boundary separation parameter, and neither of these effects were observed. Although false alarms decreased overall from the baseline to the critical task, this change did not differ by condition, and boundary separation values did not vary as a function of the instructional manipulation. Together, these findings suggest that ABST did not induce a uniformly more conservative response criterion or a global shift in response caution.

Given the clear effect of ABST on omissions, it is important to consider how this effect fits with each of the ABST theoretical frameworks. Importantly, the observed increase in omissions is not uniquely diagnostic of either account. According to the regulatory focus account, ABST induces a prevention orientation that encourages participants to withhold low-confidence responses rather than risk committing an error. We therefore preregistered the prediction that ABST would increase omissions and that such a pattern would support the regulatory focus account. However, under the task constraints used in the present study, omissions could also arise for reasons unrelated to strategic caution. Although this possibility was not considered when preregistering our hypotheses, the 1,500-ms response deadline required participants to make rapid decisions, and failures to respond before this deadline may reflect difficulty maintaining or updating information quickly enough for evidence accumulation to reach a decision threshold. Thus, omissions under deadline conditions may reflect either a strategic decision to withhold low-confidence responses or slower or less efficient processing. This interpretation aligns with prior work showing that omissions during n-back are linked to task engagement and the efficiency of information updating (Meule, 2017). Accordingly, although the ABST-related increase in omissions is consistent with the regulatory focus account, it could also reflect reduced working-memory efficiency, as predicted by the integrated process model. Because both mechanisms could plausibly produce the observed pattern under a strict response deadline, the present design cannot fully disentangle these possibilities. Future research should aim to disentangle these possibilities by manipulating response deadlines or confidence requirements to clarify whether increased omissions as a result of ABST reflect slowed evidence accumulation or deliberate response withholding.

While many studies have sought to identify a single framework that best explains ABST effects, it is also possible that the integrated process model and the regulatory focus account are not mutually exclusive. To clarify why ABST impairs older adults' performance, researchers have often contrasted predictions from these two accounts (e.g., Barber et al., 2015; Barber & Mather, 2013a; Popham & Hess, 2015; Smith et al., 2017; Thomas et al., 2020). However, this approach overlooks the possibility that both mechanisms may operate simultaneously and that their relative influence could vary across individuals and contexts. For instance, consistent with the integrated process model, ABST may elicit high levels of physiological arousal in some older adults, which in turn would result in reductions of working-memory performance. However, consistent with the regulatory focus account, for other older adults, ABST may instead trigger a stronger prevention focus, resulting in greater caution and a tendency to withhold low-confidence responses. In principle, concurrent operation of these mechanisms might be reflected in a pattern where ABST affects both the efficiency of evidence accumulation (drift rate) and response caution (boundary separation). However, we did not observe this pattern in the present data. Going forward, future research should build on the approach taken in the present study by moving beyond attempts to identify a single explanatory framework for all ABST effects and instead examining individual differences in the cognitive and motivational pathways through which ABST influences performance.

Although findings from the present study help advance our understanding of ABST effects, there are several limitations that should be noted. First, the instructional manipulation used to elicit ABST was intentionally explicit and included several elements commonly used in ABST research (e.g., framing the task as diagnostic of age-related memory decline, identifying participants as members of the stereotyped group, and referencing comparison with younger adults). This approach was chosen because the goal of the present study was to test mechanistic predictions about how ABST, once activated, alters cognitive performance. Our manipulation was therefore designed to maximize the likelihood of inducing feelings of ABST. However, this explicit wording is unlikely to occur in many real-world or clinical testing contexts, and future research is needed to examine whether similar patterns of ABST effects emerge under more subtle or naturalistic stereotype cues that better approximate everyday assessment environments.

Second, although the EZ-diffusion method (Wagenmakers et al., 2007) provided a practical and well-validated approach to estimating model parameters, this approach has limitations. For instance, it assumes that all participants begin each decision process from the same starting point, which represents the initial bias toward one response option over another. In the present study, the assumption that the starting point is midway between the two response boundaries is reasonable because match and nonmatch trials were equally likely to occur during the working-memory task. However, if ABST encouraged a more cautious response strategy that prioritized avoiding errors of commission over maximizing hits, this could have shifted the starting point toward the "nonmatch"/"new" response boundary. The EZ-diffusion method is not capable of capturing such potential shifts in response bias. Future research should use modeling techniques that allow for individual variability in starting points, providing a more comprehensive understanding of how ABST influences the underlying dynamics of decision making.

An additional limitation of the present study is that the 1,500-ms response deadline used in the working-memory tasks may have constrained participants' ability to adopt more cautious response strategies, thereby reducing the likelihood of detecting ABST-related differences in RTs or boundary separation. This design feature also limited our ability to interpret the increase in omissions observed under ABST, because failures to respond within the deadline could reflect either a strategic tendency to withhold uncertain responses (consistent with the regulatory focus account) or difficulty accumulating sufficient evidence quickly enough to reach a decision threshold (consistent with the integrated process model). While the length of this response deadline was based upon pilot testing to ensure adequate time for correct responding without the need for fast guesses, future research would benefit from systematically manipulating response deadlines to determine how time pressure modulates ABST effects on performance and help disentangle whether increased omissions under ABST reflect strategic response withholding or slowed evidence accumulation.

Finally, when designing the present study, we assumed an effect size of $d = .373$ for ABST, which reflects the meta-analytic average effect size of ABST on overall working-memory performance (Armstrong et al., 2017). However, the true effect size may be smaller, as meta-analytic estimates can be inflated by publication bias and by questionable research practices such as *p* hacking (Friese & Frankenbach, 2020). It is also possible that the effect of ABST on specific components of working-memory performance may be weaker than effects of ABST on aggregate working-memory scores. This may be particularly true when considering how ABST impacts the boundary separation parameter. In diffusion models, the drift rate is typically estimated with higher precision than the boundary separation (Ratcliff, 2013; van Ravenzwaaij & Oberauer, 2009), and estimation precision further depends on the number of trials included. Because our working-memory tasks each included only 150 trials, the DDM parameter estimations may not have been sufficiently reliable to detect ABST effects.

Despite these limitations, the present findings carry important practical implications for clinical and assessment contexts. Prior research indicates that health care and testing environments often elicit feelings of ABST among older adults (Phibbs & Hooker, 2018) and such experiences have been linked to greater physician distrust, poorer self-rated health, higher rates of hypertension, reduced engagement in preventive health care, and increased depressive symptoms (Abdou et al., 2016). ABST can also undermine older adults' subjective perceptions of their cognitive abilities, leading to lower memory self-efficacy (Bouazzaoui et al., 2016) and a greater tendency to attribute everyday memory lapses to dementia (Caughie et al., 2023). The current results suggest that in such ABST-eliciting contexts, older adults may not only demonstrate reductions in correct responding during assessments but also elevated rates of nonresponses (at least when the assessments impose strict time limits). To mitigate these adverse effects, clinicians and evaluators should consider minimizing age-related cues in test instructions and provide reassurance to help bolster older adults' confidence. Increasing clinician awareness of ABST and its potential impact is also critical, as practitioners may be unaware of its effects, especially when they themselves hold negative beliefs about aging (Parker et al., 2023).

In conclusion, the present study found that ABST was associated with reduced improvements in hit rates and increased omissions across task phases. However, each of these outcomes can be interpreted as reflecting either reduced processing efficiency or a strategic shift toward more cautious responding. Together, these results illustrate the difficulty of disentangling cognitive and motivational influences in ABST effects and emphasize the need for future research to employ designs capable of distinguishing between these underlying processes. In doing so, it will be important to consider the possibility that these two frameworks are not necessarily mutually exclusive and may instead operate together to shape older adults' performance in ABST-eliciting situations.

References

Abdou, C. M., Fingerhut, A. W., Jackson, J. S., & Wheaton, F. (2016). Healthcare stereotype threat in older adults in the health and retirement study. *American Journal of Preventive Medicine*, *50*(2), 191–198. https://doi.org/10.1016/j.amepre.2015.07.034

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

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

Adler, N. E., Epel, E. S., Castellazzo, G., & Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy, White women. *Health Psychology*, *19*(6), 586–592. https://doi.org/10.1037/0278-6133.19.6.586

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: Implications for stereotype-threat research and theory development. *Perspectives on Psychological Science*, *12*(1), 62–90. https://doi.org/10.1177/1745691616656345

Barber, S. J. (2020). The applied implications of age-based stereotype threat for older adults. *Journal of Applied Research in Memory and Cognition*, *9*(3), 274–285. https://doi.org/10.1016/j.jarmac.2020.05.002

Barber, S. J., & Lui, K. (2020). Stereotype threat and the cognitive performance of older adults. In A. K. Thomas & A. Gutchess (Eds.), *The Cambridge handbook of cognitive aging: A life course perspective* (pp. 400–416). Cambridge University Press. https://doi.org/10.1017/9781108552684.025

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

Barber, S. J., & Mather, M. (2013b). Stereotype threat can reduce older adults' memory errors. *Quarterly Journal of Experimental Psychology*, *66*(10), 1888–1895. https://doi.org/10.1080/17470218.2013.840656

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

Barber, S. J., Shoemaker, K. J., & Dotson, V. M. (2024). Ageism harms older people's health, cognition, and well-being: A selective review and policy recommendations. *Policy Insights from the Behavioral and Brain Sciences*, *11*(2), 190–198. https://doi.org/10.1177/23727322241268400

Boag, R. J., Innes, R. J., Stevenson, N., Bahg, G., Busemeyer, J. R., Cox, G. E., Donkin, C., Frank, M. J., Hawkins, G. E., Heathcote, A., Hedge, C., Lerche, V., Lilburn, S. D., Logan, G. D., Matzke, D., Miletić, S., Osth, A. F., Palmeri, T. J., Sederberg, P. B., … Forstmann, B. U. (2025). An expert guide to planning experimental tasks for evidence-accumulation modeling. *Advances in Methods and Practices in Psychological Science*, *8*(2). https://doi.org/10.1177/25152459251336127

Bouazzaoui, B., Follenfant, A., Ric, F., Fay, S., Croizet, J.-C., Atzeni, T., & Taconnat, L. (2016). Ageing-related stereotypes in memory: When the beliefs come true. *Memory*, *24*(5), 659–668. https://doi.org/10.1080/09658211.2015.1040802

Brink, T. L., Yesavage, J. A., Lum, O., Heersema, P. H., Adey, M., & Rose, T. L. (1982). Screening tests for geriatric depression. *Clinical Gerontologist*, *1*(1), 37–43. https://doi.org/10.1300/J018v01n01_06

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

Busemeyer, J. R., & Diederich, A. (2010). *Cognitive modeling*. Sage Publications.

Caughie, C., Kronenberger, O., Cobb, J., Margaris, H., McFarland, C., & Hall, S. (2023). Age-based stereotype threat and neuropsychological performance in older adults. *Aging, Neuropsychology, and Cognition*, *30*(4), 620–637. https://doi.org/10.1080/13825585.2022.2068498

Chen, Y.-T., McDonough, I. M., Faig, K. E., Norman, G. J., & Gallo, D. A. (2022). Impact of stereotype threat on brain activity during memory tasks in older adults. *NeuroImage*, *260*, Article 119413. https://doi.org/10.1016/j.neuroimage.2022.119413

Coudin, G., & Alexopoulos, T. (2010). 'Help me! I'm old!' How negative aging stereotypes create dependency among older adults. *Aging & Mental Health*, *14*(5), 516–523. https://doi.org/10.1080/13607861003713182

Fresson, M., Dardenne, B., Geurten, M., & Meulemans, T. (2017). The effect of stereotype threat on older people's clinical cognitive outcomes: Investigating the moderating role of dementia worry. *The Clinical Neuropsychologist*, *31*(8), 1306–1328. https://doi.org/10.1080/13854046.2017.1307456

Friese, M., & Frankenbach, J. (2020). p-Hacking and publication bias interact to distort meta-analytic effect size estimates. *Psychological Methods*, *25*(4), 456–471. https://doi.org/10.1037/met0000246

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

Ging-Jehli, N. R., & Ratcliff, R. (2020). Effects of aging in a task-switch paradigm with the diffusion decision model. *Psychology and Aging*, *35*(6), 850–865. https://doi.org/10.1037/pag0000562

Grimm, L. R., Markman, A. B., & Maddox, W. T. (2012). End-of-semester syndrome: How situational regulatory fit affects test performance over an academic semester. *Basic and Applied Social Psychology*, *34*(4), 376–385. https://doi.org/10.1080/01973533.2012.693427

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

Hays, R. D., Sherbourne, C. D., & Mazel, R. M. (1993). The RAND 36-Item Health Survey 1.0. *Health Economics*, *2*(3), 217–227. https://doi.org/10.1002/hec.4730020305

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), P3–P11. 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. *The Journals of Gerontology: Series B*, *64*(4), 482–486. https://doi.org/10.1093/geronb/gbp044

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

Higgins, E. T. (1998). Promotion and prevention: Regulatory focus as a motivational principle. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 30, pp. 1–46). Academic Press. https://doi.org/10.1016/S0065-2601(08)60381-0

Jang, H., Lewis, R., & Lustig, C. (2022). Opposite reactions to loss incentive by young and older adults: Insights from diffusion modeling. *Psychology and Aging*, *37*(4), 469–485. https://doi.org/10.1037/pag0000689

Lockwood, P., Jordan, C. H., & Kunda, Z. (2002). Motivation by positive or negative role models: Regulatory focus determines who will best inspire us. *Journal of Personality and Social Psychology*, *83*(4), 854–864. https://doi.org/10.1037/0022-3514.83.4.854

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

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

Meule, A. (2017). Reporting and interpreting working memory performance in n-back tasks. *Frontiers in Psychology*, *8*, Article 352. https://doi.org/10.3389/fpsyg.2017.00352

Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging*, *23*(1), 104–118. https://doi.org/10.1037/0882-7974.23.1.104

Opdebeeck, C., Yates, J. A., Kudlicka, A., & Martyr, A. (2019). What are subjective cognitive difficulties and do they matter? *Age and Ageing*, *48*(1), 122–127. https://doi.org/10.1093/ageing/afy148

Parker, G. J., Haslam, C., Stuart, J., Shum, D. H. K., & Ownsworth, T. (2023). Health practitioner beliefs regarding the impact of age-based stereotype threat on performance in the cognitive assessment of older adults. *Aging & Mental Health*, *27*(6), 1142–1155. https://doi.org/10.1080/13607863.2022.2116399

Pedersen, M. L., Alnæs, D., van der Meer, D., Fernandez-Cabello, S., Berthet, P., Dahl, A., Kjelkenes, R., Schwarz, E., Thompson, W. K., Barch, D. M., Andreassen, O. A., & Westlye, L. T. (2023). Computational modeling of the N-back task in the ABCD study: Associations of drift diffusion model parameters to polygenic scores of mental disorders and cardiometabolic diseases. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *8*(3), 290–299. https://doi.org/10.1016/j.bpsc.2022.03.012

Phibbs, S., & Hooker, K. (2018). An exploration of factors associated with ageist stereotype threat in a medical setting. *The Journals of Gerontology: Series B*, *73*(7), 1160–1165. https://doi.org/10.1093/geronb/gbx034

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

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Ratcliff, R. (2013). Parameter variability and distributional assumptions in the diffusion model. *Psychological Review*, *120*(1), 281–292. https://doi.org/10.1037/a0030775

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the lexical-decision task. *Psychology and Aging*, *19*(2), 278–289. https://doi.org/10.1037/0882-7974.19.2.278

Ratcliff, R., Thapar, A., & McKoon, G. (2007). Application of the diffusion model to two-choice tasks for adults 75–90 years old. *Psychology and Aging*, *22*(1), 56–66. https://doi.org/10.1037/0882-7974.22.1.56

Régner, I., & Huguet, P. (2025). Age-based stereotype threat effects: From the laboratory to the clinical setting. *Cortex*, *183*, 261–273. https://doi.org/10.1016/j.cortex.2024.12.006

Rossi-Arnaud, C., Spataro, P., & Geraci, L. (2018). Effects of stereotype threat and prior task success on older adults' eyewitness memory. *Journal of Applied Research in Memory and Cognition*, *7*(3), 422–431. https://doi.org/10.1037/h0101822

Salthouse, T. A. (2004). What and when of cognitive aging. *Current Directions in Psychological Science*, *13*(4), 140–144. https://doi.org/10.1111/j.0963-7214.2004.00293.x

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

Sewell, D. K., Lilburn, S. D., & Smith, P. L. (2016). Object selection costs in visual working memory: A diffusion model analysis of the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(11), 1673–1693. https://doi.org/10.1037/a0040213

Shepherdson, P., Oberauer, K., & Souza, A. S. (2018). Working memory load and the retro-cue effect: A diffusion model account. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(2), 286–310. https://doi.org/10.1037/xhp0000448

Smith, A. M., Gallo, D. A., Barber, S. J., Maddox, K. B., & Thomas, A. K. (2017). Stereotypes, warnings, and identity-related variables influence older adults' susceptibility to associative false memory errors. *The Gerontologist*, *57*(suppl_2), S206–S215. https://doi.org/10.1093/geront/gnx057

Spaniol, J., Madden, D. J., & Voss, A. (2006). A diffusion model analysis of adult age differences in episodic and semantic long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 101–117. https://doi.org/10.1037/0278-7393.32.1.101

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

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

Swift, H. J., Abrams, D., & Marques, S. (2013). Threat or boost? Social comparison affects older people's performance differently depending on task domain. *The Journals of Gerontology: Series B*, *68*(1), 23–30. https://doi.org/10.1093/geronb/gbs044

Theisen, M., Lerche, V., von Krause, M., & Voss, A. (2021). Age differences in diffusion model parameters: A meta-analysis. *Psychological Research*, *85*(5), 2012–2021. https://doi.org/10.1007/s00426-020-01371-8

Thomas, A. K., Smith, A. M., & Mazerolle, M. (2020). The unexpected relationship between retrieval demands and memory performance when older adults are faced with age-related stereotypes. *The Journals of Gerontology: Series B*, *75*(2), 241–250. https://doi.org/10.1093/geronb/gby031

van Ravenzwaaij, D., & Oberauer, K. (2009). How to use the diffusion model: Parameter recovery of three methods: EZ, fast-dm, and DMAT. *Journal of Mathematical Psychology*, *53*(6), 463–473. https://doi.org/10.1016/j.jmp.2009.09.004

Verhaeghen, P., Marcoen, A., & Goossens, L. (1993). Facts and fiction about memory aging: A quantitative integration of research findings. *Journal of Gerontology*, *48*(4), P157–P171. https://doi.org/10.1093/geronj/48.4.P157

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, *32*(7), 1206–1220. https://doi.org/10.3758/BF03196893

Wagenmakers, E.-J., van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*(1), 3–22. https://doi.org/10.3758/BF03194023

Wong, J. T., & Gallo, D. A. (2016). Stereotype threat reduces false recognition when older adults are forewarned. *Memory*, *24*(5), 650–658. https://doi.org/10.1080/09658211.2015.1036885

Zhang, B., Lin, Y., Gao, Q., Zawisza, M., Kang, Q., & Chen, X. (2017). Effects of aging stereotype threat on working self-concepts: An event-related potentials approach. *Frontiers in Aging Neuroscience*, *9*, Article 223. https://doi.org/10.3389/fnagi.2017.00223

Received December 1, 2025
Revision received March 6, 2026
Accepted April 29, 2026 ■