



Published in final edited form as:

*Neuroimage*. 2022 October 15; 260: 119413. doi:10.1016/j.neuroimage.2022.119413.

## Impact of stereotype threat on brain activity during memory tasks in older adults

Yung-Tsen Chen<sup>a</sup>, Ian M. McDonough<sup>b</sup>, Kelly E. Faig<sup>c</sup>, Greg J. Norman<sup>a</sup>, David A. Gallo<sup>a,\*</sup>

<sup>a</sup>Department of Psychology, University of Chicago, 5848 S. University Ave., Chicago, IL, 60637, USA

<sup>b</sup>Department of Psychology, The University of Alabama, 505 Hackberry Lane, BOX 870348, Tuscaloosa, AL 35487, USA

<sup>c</sup>Psychology Department, Hamilton College, 198 College Hill Road, Clinton, NY 13323, USA

### Abstract

We report the first neuroimaging experiment to investigate the impact of explicitly activating aging stereotypes (i.e., stereotype threat) on brain activity during cognitive tasks. Cognitively normal older adults read about aging stereotypes or a control passage prior to taking episodic memory, working memory, and a non-demanding control task during fMRI. At the group level, stereotype activation did not impact cognitive performance or measures sensitive to stress and anxiety (physiological or self-report), but like prior work, highly educated and retired adults exhibited greater stereotype effects on episodic memory. At the neural level, stereotype activation did not impact brain activity in executive control or emotional regulation regions previously linked to stereotype threat effects in younger adults, suggesting that stereotype threat operates differently in older adults. Instead, on each task, the stereotype group showed more brain activity than the control group in parietal midline regions (e.g., precuneus, posterior cingulate). Although activity in these regions can arise from many processes, they have previously been associated with self-referential thinking and error-prevention focus, and in our study, brain activity in these regions was associated with slower responses and lower false alarm errors on the episodic memory task. Collectively, these findings are more consistent with the regulatory fit hypothesis than an executive control interference hypothesis of stereotype threat effects in older adults, whereby older adults adopt an error-prevention mindset in response to explicit stereotype threat.

### Keywords

Aging; Episodic memory; Functional magnetic resonance imaging; Working memory; Stereotype threat

---

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

\*Corresponding author. [dgallo@uchicago.edu](mailto:dgallo@uchicago.edu) (D.A. Gallo).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.119413.

## 1. Introduction

Explicitly activating negative stereotypes about aging can impair cognitive performance in older adults, especially during memory tasks. This effect has been attributed to the negative impact of stereotype threat, whereby activating negative stereotypes can impair an individual's performance in stereotyped domains if they feel the stereotype applies to them. For example, stereotype activation can decrease accuracy on short-term working memory tasks and long-term episodic memory tasks in older adults (Hess et al., 2003; Kang and Chasteen, 2009; Mazerolle et al., 2012). Such age-related impairments have been confirmed in several meta-analyses with a moderate effect size (Armstrong et al., 2017; Lamont et al., 2015; Meisner, 2012). Although these effects have been replicated, the underlying mechanisms remain in question. Better understanding the mechanisms of aging stereotype effects can allow researchers and clinicians to design better measures of neuropsychological decline, predict when such effects might occur, and develop interventions to target those mechanisms. Without such knowledge, stereotype activation can confound performance on cognitive tests, including those used to detect Alzheimer's disease and related dementias, thereby leading to misdiagnoses of cognitive impairment (Haslam et al., 2012).

One explanation of stereotype effects on cognition is the executive-control interference hypothesis, which is based primarily on stereotype threat effects in stigmatized groups of younger adults (e.g., math performance in female younger adults; Pennington et al., 2016; Spencer et al., 2016). This hypothesis claims that activating negative stereotypes increases negative emotions, self-reflection, rumination, and anxiety—all of which must be suppressed with cognitive effort (e.g., attentional control or working memory resources). This additional effort creates a competition for executive-control resources that might otherwise be devoted to the cognitive task (Beilock et al., 2007; Schmader et al., 2008). The executive-control interference hypothesis can readily explain the negative impact of aging stereotypes on cognitive tasks in older adults, especially given that aging tends to reduce executive control and increase susceptibility to interference more generally (Hasher and Zacks, 1988).

Although the executive-control interference hypothesis can explain the negative impact of stereotype activation on performance, activating aging stereotypes has not always been found to reduce executive control resources or working memory in older adults (for reviews, see Barber, 2017; Barber and Mather, 2014). Such findings have led to challenging this hypothesis as a primary explanation of aging-related stereotype effects on cognition. An alternative hypothesis to explain the impact of aging stereotypes derives from the regulatory fit hypothesis (Barber and Mather, 2013a, 2013b). According to this hypothesis, explicitly activating aging stereotypes does not necessarily cause negative emotions and anxiety that need to be suppressed in older adults, but instead, it motivates participants to change cognitive strategies by adopting either a promotion-focus or a prevention-focus mindset (Higgins, 1998, see Seibt and Forster, 2004). When negative stereotypes are activated under typical cognitive task conditions, this hypothesis assumes that older adults tend to adopt a prevention-focus mindset. This shift in strategy leads to more conservative responding that can reduce errors of commission (e.g., reduced false alarms), but also can undermine gains on cognitive tasks and increase errors of omission (e.g., reduced hits, see Barber, 2017). The ultimate impact of this strategy shift on performance should depend on the

task, the interpretation of the instructions, and the types of strategies that lead to optimal (or suboptimal) performance. For example, activating aging stereotypes can either increase or decrease working memory performance in older adults depending on the task's payoff structure, which was argued to be more consistent with the regulatory fit hypothesis than executive-control interference (Barber and Mather, 2013a).

Adjudicating between the executive-control interference and regulatory fit hypotheses has been difficult using behavioral performance alone. Although the regulatory fit hypothesis can explain how some situations lead to improved task performance, both hypotheses propose that stereotype threat activation under typical cognitive task conditions can impair performance, which are the most typical findings in the aging literature. Moreover, the impact of explicit aging stereotypes on memory performance is irregular across studies in the literature (e.g., Hess et al., 2009), potentially owing to different task demands or individual difference variables that may interact with stereotype activation. For example, older adults sometimes demonstrate elevated false alarms on episodic memory tasks when aging stereotypes are explicitly activated during encoding (Wong and Gallo, 2019) or retrieval (Smith et al., 2017; Thomas and Dubious, 2011), but these effects can be eliminated or reversed under task conditions believed to encourage an error-prevention focus (Barber and Mather, 2013b; Thomas et al., 2020; Wong and Gallo, 2016). Unfortunately, predicting which task conditions might elicit an error-prevention focus has proved difficult and research in this area would benefit from an independent measure of underlying processes.

One way to advance our understanding of these effects would be to supplement behavioral measures with other measures, such as brain activity, which could provide additional evidence to further delineate the relevant hypotheses. To this end, the current experiment used neuroimaging to investigate the impact of aging stereotype activation during cognitive tasks in older adults. Our goal was to identify the impact of stereotype activation on brain activity (fMRI) and measures sensitive to stress response (high-frequency heart rate variability; HF HRV) and anxiety (self-reported), and to use these measures along with cognitive performance to test between alternative hypotheses of aging stereotype effects on cognition. We are unaware of any prior neuroimaging studies investigating the explicit activation of aging stereotypes during cognitive tasks in older adults. Nevertheless, as reviewed next, prior literature predicts different patterns of brain activity and physiological response associated with the executive-control interference hypothesis compared to the regulatory fit hypotheses, leading to differing predictions for stereotype threat effects on these measures in older adults.

There have been two published fMRI studies of stereotype threat in younger adults (Krendl et al., 2008; Wraga et al., 2007), and both studies found results that support the executive-control interference hypothesis. Specifically, activating stereotypes impaired cognitive performance and increased activity in brain regions involved in emotional regulation and executive control (i.e., ventral anterior cingulate cortex in both studies, amygdala, medial frontal gyrus, and ventrolateral prefrontal cortex in Wraga et al., 2007). Elevated activity in these regions suggests that participants had negative emotional reactions when confronted with negative stereotypes, potentially causing interference with executive control processes that otherwise would be devoted to the cognitive tasks. EEG studies of stereotype threat in

younger adults also have been consistent with this hypothesis (see Derks et al., 2008; Forbes and Leitner, 2014; Mangels et al., 2012). To the extent that negative aging stereotypes yield similar reactions in older adults, we predicted similar patterns of brain activity in response to stereotype threat in older adults in our study. The executive-control interference hypothesis also predicts a negative stress response associated with stereotype activation and resulting need for emotional regulation, which would be reflected in ratings of anxiety and heart rate variability.

The regulatory fit hypothesis does not predict these same fMRI and physiological effects because it does not assume the involvement of negative emotions or emotional regulation in response to stereotype threat in older adults. This hypothesis instead argues that older adults might view the negative stereotype passage as a threat to their self-image, thereby motivating them to avoid losses and prevent errors (a prevention-focus). Two previous fMRI studies have investigated the neural correlates of prevention focus as specified by the regulatory fit framework (Johnson et al., 2006; Mitchell et al., 2009). These fMRI studies did not involve stereotype activation, but instead required participants to think about avoiding personal losses (prevention focus), compared to thinking about improving gains (a promotion focus), and a non-self-referential control task. In younger adults, Johnson et al. (2006) found that both tasks activated medial frontal regions (medial frontal gyrus/anterior cingulate) and posterior areas (posterior cingulate cortex [PCC], precuneus) previously linked to self-referential processing, with a prevention focus specifically associated with greater activity in PCC and adjacent regions compared to promotion focus and the control task. The link between prevention focus and medial posterior regions was replicated in Mitchell et al. (2009), and older adults also demonstrated greater PCC activity during prevention focus compared to the control task in that study. Although these studies did not manipulate stereotype threat, a study by Colton et al. (2013) found increases in the PCC and mid cingulate cortex (MCC) when older adults were rating words related to aging stereotypes (compared to neutral words) for self-relevance, implicating these regions in the processing of aging stereotypes. Together, these findings predict that older adults should have greater PCC activity in the stereotype group compared to the control group to the extent that negative stereotypes elicit self-referential thought and a prevention focus. Of course, the executive control interference hypothesis also involves self-referential thought, but the two fMRI studies with younger adults described above did not find activity in PCC or surrounding regions in association with negative stereotypes, suggesting that posterior medial regions are not critical for the emotional regulation processes assumed by that hypothesis.

## 2. Current study

In the current study we experimentally manipulated the activation of aging stereotypes just prior to measuring brain activity and cognitive performance on memory tasks. In the first session, prior to any manipulation, all participants took baseline memory tests and later were scheduled for an MRI in which they took the same kinds of tests with new items. Immediately before the MRI session, participants were randomly assigned to a group that read a passage stating that the purpose of the study was to better understand the negative impact of cognitive aging and Alzheimer's disease on memory (stereotype group)

or to help understand individual differences in human brain functioning, with no mention of aging (control group). Heart-rate variability and self-reported anxiety were assessed before and after this stereotype manipulation. Immediately thereafter, during fMRI scanning, participants received both a working memory task (N-back with a load of 2 items) and an episodic memory task (memorizing and recognizing categorized words), as well as a non-demanding control task (reading a number countdown). The episodic memory task was modeled after previous studies that used categorized words and found reduced false alarms when stereotypes were activated (Barber and Mather, 2013b; Wong and Gallo, 2016). The N-back task was selected because it relies heavily on working memory and reliably activates executive control regions in prefrontal cortex during fMRI (Nee et al., 2013; Owen et al., 2005).

If stereotype threat induces executive-control interference in older adults, then we would expect elevated activity in regions previously associated with emotional regulation and executive control (i.e., ventral anterior cingulate cortex, amygdala, lateral prefrontal cortex) on both memory tasks. We also would expect a stress response during the threat manipulation, which in this context should be reflected in increased anxiety and associated decreases in high-frequency heart rate variability (see Williams et al., 2019). By contrast, if stereotype threat induces a shift towards a prevention focus, then we would not expect elevated activity in these regions or this kind of stress response. Instead, the studies reviewed previously predict that we would find elevated activity in medial posterior regions associated with self-referential thought and prevention focus (i.e., posterior cingulate cortex) while older adults are taking the cognitive tasks.

In addition to testing between these two hypotheses, we also explored individual differences that might moderate the effects of stereotype activation in ways that are still poorly understood. Smith et al. (2017) argued that being retired and/or highly educated may increase one's concern about cognitive decline with aging that, in turn, increases their sensitivity to stereotype activation (see also Andreoletti & Lachman, 2004; Hess et al., 2009; Kang and Chasteen, 2009; for contradictory evidence, see Armstrong et al., 2017). We therefore included these variables in our analyses.

## 2.1. Participants

One hundred and two cognitively normal older adults from throughout the Chicago area were tested in the laboratory (pre-fMRI session). These participants were recruited via advertisements for a study on individual differences in MRI brain activity, to avoid revealing the aging nature of the study (and hence activating stereotype threat prior to our manipulation). Recruitment targeted the age range from 60 to 70 because participants in this range (i.e., early older adulthood) have been found to be more sensitive to explicit aging stereotype manipulations than those in late older adulthood (i.e., over 70, see Hess et al., 2009). All participants reported normal or corrected-to-normal vision, no hearing problems, no history of neurological or psychiatric condition associated with cognitive decline or chronically poor health (e.g., Alzheimer's disease or MCI, Parkinson's disease, stroke, heart attack), no history of alcohol/narcotics abuse or recent head trauma, reported English as a primary language (learned by age 6), and were screened for standard MRI compliance

(e.g., right-handed, as confirmed by Edinburgh Handedness Inventory, Oldfield, 1971, no uncontrolled blood pressure, no MRI incompatible metal implants or devices). Older adults were paid \$50 for a lab visit and \$100 for a subsequent fMRI session. We also tested a group of younger adults (data not included in this report). Out of the 102 older adults tested in the lab, 73 came back for the fMRI session (14 were not MRI eligible, 13 performed insufficiently on the pre-fMRI cognitive tasks to continue, and 2 did not receive fMRI due to scheduling difficulty). Out of the 73 participants receiving fMRI, data from 1 participant was lost due to technical difficulties, and data from 3 older adults (blocked design) were discarded due to excessive motion, yielding 69 participants.

Sociodemographic and behavioral data are reported based on the sample of 69 who completed both sessions and with analyzable images (see Table 1). Older adults in the stereotype group did not differ from the control group in global cognition as measured by the MoCA (Nasreddine et al., 2005) or Shipley verbal and analytical subscales (Shipley et al., 2009), and they also did not differ in responses to questionnaires on trait anxiety (Spielberger, 2010) or tendencies to worry (Meyer et al., 1990). However, compared to the control group, the stereotype group was slightly older, reported higher ratings of memory ability on the Memory Controllability Inventory questionnaire (Lachman et al., 1995), reported lower rates of rumination on the Rumination Reflection Questionnaire (Trapnell and Campbell, 1999), and lower reports of perceived stereotype threat of aging in their general lives (Kang and Chasteen, 2009). Note that this last questionnaire included 4 questions assessing perceptions of stereotype threat in general (i.e., during their daily life), and one question about perceived threat specific to the experimental session. The group difference on the experiment-specific question was not significant, and average ratings in both groups indicated they did not think the experimenter expected them to do poorly because of their age (average ratings fell between “strongly disagree” and “disagree”).

**Human Ethics Statement:** All participants gave written informed consent to participate in the research, and all procedures were in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) and approved by the University of Chicago Social and Behavioral Sciences IRB.

## 2.2. Procedures

Each older adult was tested on two occasions. The first visit occurred 2 to 4 weeks before fMRI scanning and assessed baseline performance on the memory tests. After completing the cognitive tasks, participants took the MoCA and Shipley scales and only participants who scored 18 or higher on the MoCA and had accuracy of at least 10% (hits – FAs) on the episodic and working memory tests were recruited to come back for the fMRI session (to ensure they could perform the tasks). At the end of the first session, we administered questionnaires to assess rumination tendencies (Trapnell and Campbell, 1999), worry tendencies (Meyer et al., 1990), and trait anxiety (Spielberger, 2010), and to disguise the nature of the study, we also administered scales on authority compliance (Gudjonsson, 1989) and paranormal beliefs (Thalbourne and Delin, 1993), there were no group differences on these last two measures (both  $p$ 's > 0.10). No age-related questions or age-sensitive terms were used during this first session, to avoid activating aging stereotypes.

During the second session, participants were connected with electrodes to assess high-frequency heart rate variability, a measure of parasympathetic activity. After recording five-minutes while resting, and taking a questionnaire on their current state anxiety (Spielberger, 2010), participants were randomly assigned to be explicitly exposed to either negative stereotypes about aging memory (stereotype group) or an age-neutral framing of the task (control group). Participants read one of two scientific passages describing the nature of the study (see Appendix). The passage either indicated that the current study was investigating tasks sensitive to cognitive impairment associated with normal aging and Alzheimer's disease (stereotype group) or that the study was investigating tasks sensitive to individual differences in brain function (control group). Following this procedure, the electrodes were removed and participants entered the MRI scanner to take the same memory tasks as in the baseline session, but with a different set of stimuli (approximately 20 to 30 min following stereotype manipulation). After scanning, participants completed a second assessment of their state anxiety, this time modified to refer to the anxiety they experienced while in the scanner, as well as the Memory Controllability Inventory (Lachman et al., 1995) and Perceived Stereotype Threat (Kang and Chasteen, 2009; see Table 1). These questionnaires were given at the end of the second session because they explicitly referenced memory and stereotypes, and we did not want them to activate stereotype threat during our experimental manipulation. At the end of the experiment, participants also received a manipulation-check question asking if they recalled the experimenter telling them that the study was about cognitive decline with normal aging or Alzheimer's disease, and if so, how they thought this impacted them.

### 2.3. Cognitive tests

The same cognitive tasks were given during baseline and in the MRI scanner, but with different items in each session. The cognitive tasks were given in two phases. The first phase consisted of three interleaved tasks using a blocked design (episodic memory encoding, N-back, and a low-level numerical countdown task, Fig. 1). The second phase consisted of an event-related recognition memory test for the episodic memory items that were previously encoded. For the blocked design, there were 6 blocks for each task, divided across 2 functional runs. For all three tasks, each block contained 11 trials (1.5 presentation duration, 250 ms ISI), matched for number of button presses and visual complexity. For the episodic memory encoding task, participants were asked to memorize the items. In total, participants encoded 4 words from each of 12 categories (e.g., furniture: table, desk, sofa, dresser), drawing from the categories eliciting the highest false recognition in Gallo (2004, Experiment 2). Items were presented along with their category labels, and each encoding block included 4 items from 2 categories (intermixed), along with three intermixed "press button" trials as an attentional control and also to match the number of button presses across all block types. The working memory test consisted of a 2-back task in which 4 letters (A, B, C or D) were randomly presented over 11 trials, and participants were asked to indicate by button press when the currently presented letter was the same as the one presented 2 trials previously (three targets per block). The countdown task presented the numbers 7, 6, 5, 4, 3, 2, 1 and 0 in order, along with 3 "press button" trials intermixed. Participants were asked to clear their mind during the countdown and to press the button when prompted. For the working memory and countdown tasks, the letters and numbers were presented as character

strings to match the visual complexity of the display with the episodic memory encoding blocks. The memory encoding and 2-back blocks alternated with countdown blocks always in between.

Following these blocks, recognition memory was tested, also in the scanner. The test comprised 144 items: 4 studied items from each of the 12 studied categories (targets), 4 nonstudied items from each of the 12 studied categories (related lures), and 4 items from 12 nonstudied categories (unrelated lures). Test items were presented along with their category labels, as during the study phase. Each test item was presented for 2.5 s, separated by a fixation that was jittered between 0 and 10 s (average ISI of 1.5 s). The test was administered across three functional runs. Participants were told that the test would contain studied items and nonstudied items that were either related to studied categories or not, with examples from a category that was not used in the tasks (1 studied item, 1 related lure, and 1 unrelated lure). Participants were told to press “7” with their index finger for new word pairs (including both related and unrelated lures), and press “8” with their middle finger for old words pairs (the ones they studied during encoding phase). Because participants were aware that some related items needed to be rejected, this aspect of the instructions promoted error-prevention.

Each task resulted in several performance indices including hit rates, false alarm rates, and a combined accuracy score (hits – false alarms). Because the episodic memory task included related and unrelated lures, this resulted in two types of memory accuracy scores (old/related accuracy and old/unrelated accuracy). Two primary sets of analyses were conducted to investigate the impact of stereotype group on behavior. First, a 2 (Group: Stereotype vs. Control)  $\times$  2 (Session: Lab vs. fMRI) repeated measure ANOVA was conducted separately for accuracy scores in the episodic and working memory tasks. The second analysis was based on prior work suggesting individual differences may mask stereotype activation effects on episodic memory (e.g., Smith et al., 2017). For this analysis we investigated a three-way interaction using group (stereotype, control), retirement status (retired, working) and years of education as predictors in an ANCOVA with the two accuracy scores as the dependent variables. Significant findings were followed by separate tests for hits and false alarms to better understand the nature of the accuracy effects.

#### 2.4. Physiological data acquisition and design

Parasympathetic cardiac control was measured with high-frequency heart rate variability (HF HRV) derived from the ECG. HF HRV is the fluctuation in heart rate between beats within the respiratory frequency band (0.12 – 0.40 Hz) that has been demonstrated to be an index of parasympathetic cardiac control (Berntson et al., 2017). HF HRV is represented as the natural log of the heart period variance in the respiratory band (in  $\text{ms}^2$ ). The ECG was obtained using a Bionex system (Mindware Technologies LTD, Gahanna, OH). Mindware software was used to derive HF HRV by spectral analysis of the interbeat interval series from the ECG.



## 2.5. fMRI acquisition

Images were acquired using a 3-T Philips Achieva scanner at the University of Chicago MRI Research Center. Functional scans were acquired using a  $T_2^*$ -weighted echo planar imaging sequence (repetition time [TR] = 2 s, echo time [TE] = 25 ms, field of view [FOV] = 192 mm; flip angle =  $77^\circ$ , matrix size =  $64 \times 63$  mm, in-plane resolution =  $3 \times 3$  mm<sup>2</sup>). For whole-brain coverage, 34 interleaved slices (4 mm thickness, 0.5 mm skip between slices) were acquired. During each of the 5 functional runs, we applied z-shimming on 4 slices across orbital frontal region along superior/inferior direction for compensation gradient to regain signal loss due to nasal cavity artifact (Du et al., 2007). Structural scans were acquired last using high resolution  $T_1$  - weighted structural Turbo Field Echo (TR = 8 ms, TE = 3.8 ms, flip angle =  $8^\circ$ , FOV = 224 mm, matrix =  $224 \times 224$  mm<sup>2</sup>, in-plane resolution =  $1 \times 1$  mm<sup>2</sup>).

## 2.6. fMRI data processing and analysis

Our MRI tasks had two phases: a blocked task phase which alternated between the episodic memory encoding, working memory, and countdown tasks, followed by an event-related memory test for the episodic memory items. Our stereotype manipulation preceded both phases, so an impact of stereotype activation on participants' overall approach to the various tasks – or their mindset during the experimental session – should have been present in each phase. Because blocked designs tend to have more power to detect group differences in performance compared to event-related designs, these were our primary focus and we only report MRI methods and analyses of the blocked design in the sections that follow. Additional methods and analyses for the event-related recognition task are described in the Supplemental Materials.

Preprocessing and data analysis were conducted using SPM12 (Wellcome Trust centre for Neuroimaging, London) implemented in MATLAB R2015a (The Mathworks Inc., Natick, MA, USA). Standard preprocessing was performed on the functional data, including using fieldmaps to unwarped the images, realigning the time series using a least squares approach and a six-parameter (rigid body) spatial transformation aligning to the first image in the series, anatomical coregistration, segmentation of the anatomical scan into gray matter, white matter, and cerebrospinal fluid, and spatial smoothing (using a 6-mm full-width half-maximum isotropic Gaussian kernel). Normalization to the Montreal Neurological Institute (MNI space) template (resampling at 2 mm cubic voxels) was implemented for group analyses.

For each participant, the BOLD response was modeled using an unbiased whole-brain approach under the assumptions of the general linear model (GLM) (Friston et al., 1995). In these analyses, blocked event types of interest and covariates of no interest (a mean for the functional run, a linear trend for the functional run, and six movement parameters derived from realignment) were used to compute parameter estimates at each voxel. A high-pass filter of 66 s was used to remove low-frequency drifts. Second level analyses were conducted using a 2 (Group: stereotype, control) by 3 (Task: episodic encoding, working memory, countdown) mixed factorial design. This ANOVA yielded overall task-related differences in brain activity (main effect of task: encoding vs. working memory vs.

countdown), overall stereotype-related differences in brain activity (main effect of group: stereotype vs. control), and any differences in the effect of stereotype on brain activity associated with these tasks (interaction between task and group). For completeness, the countdown task also was used as a low-difficulty control task to identify brain activity associated with the memory tasks, and we report contrasts for episodic encoding > countdown and working memory > countdown. All the results were reported at  $p < .05$ , corrected for multiple comparisons using Monte Carlo simulations via 3dClustSim on AFNI.

In addition to whole-brain analyses, region of interest (ROI) analyses were conducted. Anatomical ROIs were selected based on the literature to further characterize the BOLD response during the different cognitive tasks and possible group differences. ROIs linked to the executive-control interference hypothesis were based on prior stereotype threat studies in younger adults (Krendl et al., 2008; Wraga et al., 2007), and included bilateral amygdala, ACC, MFG, and left IFG/BA 47. ROIs linked to the regulatory fit hypothesis were based on prior studies of prevention focus (Johnson et al., 2006; Mitchell et al., 2009) or processing aging stereotype words (Colton et al., 2013) and included MCC and PCC. ROIs were based on the neuromorphometric atlas in SPM12 and contrast estimates were extracted using SPM12 tools. Contrast estimates were then entered into two different Group  $\times$  Condition MANOVAs to separately test the executive-control interference hypothesis and the regulatory fit hypothesis.

We conducted brain-behavior correlations using partial least squares regression (PLS-R) via the ExPosition package in R (Beaton et al., 2014). This method was chosen for several reasons. First, behavioral performance measures often are correlated with one another as is true for brain activity across regions and PLS techniques capitalize on the shared variance across factors to explain most of the covariance in the data. Second, the multivariate nature of this technique allows a simultaneous estimation of many measures, including multiple brain regions and multiple behavioral measures. Because the covariance in the variables was analyzed together in a single analysis, no corrections for multiple comparisons were needed. The brain and behavioral variables of interest were used to create two matrices: one that represented the brain measures for each participant and one that represented the behavioral measures. The cross product of these two matrices were decomposed into mutually orthogonal latent variables using singular value decomposition. The latent variable scores represented the weights of the brain factors that contributed to higher or lower performance for each behavioral measure. Pearson correlations between the latent variable scores from each X and Y matrix were used to determine significance and effect size of each resulting factor.

### 3. Results

#### 3.1. Manipulation check

Of the 69 participants in the study, all 33 participants in the stereotype group remembered reading the aging stereotype passage, whereas only 7 of 36 (19%) participants in the control group reported thinking that the study was about cognitive decline with aging. Each of these 7 participants reported they did not get this message from the researcher or from the passage but guessed the purpose of the study was related to aging or memory through the self-report

questionnaires that were given at the very end of the study before the manipulation check. None of the control participants reported thinking about aging stereotypes before or during the fMRI scanning. These results suggest that our procedures successfully activated aging stereotypes in the stereotype group only.

### 3.2. Pre-Post changes in stress response and anxiety by group

Pre and post changes in parasympathetic cardiac response and anxiety were measured as a function of the stereotype activation manipulation with resting HF HRV and self-reported state anxiety, respectively (see Table 1). Higher resting HF HRV is associated with more optimal frontal functioning and cognitive control in response to challenging tasks (Colzato and Steenbergen, 2017), whereas lower resting HF HRV tends to be associated with suboptimal stress responses (Porges, 2001; Thayer et al., 2009), and lower HRV has been associated with the negative impact of stereotype threat on performance in younger adults (Williams et al., 2019). Three time points were measured for HF HRV: pre-passage, during the passage, and post-passage. Analyses on HF HRV yielded a main effect of time ( $F(2, 114) = 7.99, p < .001$ ), reflecting an increase in HF HRV in both groups with no effect of group and no interaction ( $ps > 0.87$ ). Pre and post measures of state anxiety also were gathered. This analysis yielded a main effect time ( $F(1, 66) = 79.38, p < .001$ ), reflecting an increase in state anxiety in both groups with no effect of group and no interaction ( $ps > 0.71$ ). These data are inconsistent with the idea that the stereotype passage would make participants feel more anxious or threatened than the control passage, which was predicted by the executive-control interference hypothesis but not the regulatory fit hypothesis (which assumes strategy shifts, but not necessarily anxiety or stress in response to aging stereotypes). Instead, these findings suggest that reading a passage about the study's goals in preparation for taking the tasks during MRI elevated anxiety, with no group differences in this anticipatory response.

### 3.3. Cognitive tasks

Participants performed well on the cognitive tasks and could distinguish between targets and lures effectively (see Table 2). At the group level, there were minimal differences in performance on either task that could be attributed to stereotype activation. For episodic memory, a  $2$  (Group: Stereotype vs. Control)  $\times$   $2$  (Session: Lab vs. fMRI) repeated measure ANOVA was conducted on each of the two measures of accuracy, revealing a significant group effect on old/related accuracy ( $F(1,67) = 4.88, p = .031$ ). This effect indicates that the stereotype group outperformed the control group, but this effect did not interact between the sessions, providing no evidence that stereotype activation (in the second session) impacted performance. No other group-level effects were found on episodic memory or working memory performance (all  $ps > 0.22$ ). For completeness, Table 2 also report separate group comparisons of each hit rate, false alarm rate, and accuracy score, as well as a corrected false recognition score on the episodic memory task (false alarms to related lures minus unrelated lures), which assesses the impact of category relatedness on false alarms independent from response bias effects that should impact all lure types.

Following prior work suggesting individual differences may mask stereotype activation effects in episodic memory (e.g., Smith et al., 2017), we investigated whether retirement

status (retired, working) and years of education may moderate the impact of stereotype activation on episodic and working memory accuracy. For episodic memory, this analysis revealed that stereotype activation increased old/new accuracy in highly educated and retired individuals ( $p < .019$ ). This result was driven by reduced false alarms to new items, potentially reflecting a switch to more conservative category-based monitoring and responding in the stereotype group, which could increase old/new accuracy in the current task (see Supplemental Material; Fig. S1). Although prior work has been mixed (see Smith et al., 2017), the current result is consistent with two prior studies that also tested episodic memory for associated words, which found that stereotype activation reduced false alarms under task procedures that encouraged error-prevention as in the current task (e.g., warnings against errors, Barber and Mather, 2013b; Wong and Gallo, 2016), consistent with the regulatory fit hypothesis. For working memory, although the three-way interaction was in the same general direction, none of the effects reached significance ( $ps > 0.61$ ).

### 3.4. fMRI whole brain univariate

As discussed in the methods section, the stereotype manipulation preceded the MRI tasks, and so any general impact of stereotype threat on participants' mindset when taking the various tasks (and associated differences in brain activity) should have been observed across all our task phases. Because blocked designs tend to be more powerful than event-related designs, these were our primary focus and we only report analyses of the blocked design in the sections that follow. Nevertheless, analyses of brain activity during the event-related phase supported the key results, and these are reported in the Supplementary Material.

On the blocked MRI data, we conducted a 2 (Group: stereotype, control)  $\times$  3 (Task: episodic encoding, 2-back, countdown) factorial ANOVA to identify overall group differences between the stereotype and the control groups and Group  $\times$  Task interactions on brain activity. With respect to task-related regions, these analyses revealed robust patterns of brain activity that have previously been associated with these kinds of cognitive tasks (Fig. 2). To further investigate these task-specific effects, we conducted posthoc contrasts for each task relative to the countdown task (Table S1). For episodic memory encoding, brain activity was lateralized to the left hemisphere, including left temporal cortex, PFC, and lateral parietal cortex - both typically associated with episodic encoding using verbal materials with visual presentation (Spaniol et al., 2009; Kelley et al., 1998). No regions showed the opposite pattern (countdown  $>$  encoding) at the set threshold. For working memory, brain activity was found in bilateral insula, PFC, and inferior parietal cortex - all commonly found in N-back tasks (Owen et al., 2005; Nee et al., 2013). No regions showed the opposite pattern (countdown  $>$  2-back) at the set threshold.

With respect to stereotype and control group differences, no main effects of group were found but a significant group  $\times$  task interaction was observed in three clusters: right dorsal PCC, right precuneus, and right postcentral sulcus (see Fig. 3 and Table 3). Follow-up analyses were conducted to determine the direction of these interaction effects and to compare the size of the effect across tasks. These analyses indicated that the interaction in each cluster was due to the stereotype group exhibiting higher activity in the memory encoding task than the control group ( $t(67)=4.16$ ,  $SE = 0.13$ ,  $p < .001$ ,  $t(67)=3.17$ ,  $SE =$

0.21,  $p = .002$ , and  $t(67) = 3.11$ ,  $SE = 0.16$ ,  $p = .003$ ), whereas smaller group differences were found between the two groups in the 2-back task ( $t(67) = 2.64$ ,  $SE = 0.13$ ,  $p = .01$ ,  $t(67) = 1.96$ ,  $SE = 0.21$ ,  $p = .06$ , and  $t(67) = 2.52$ ,  $SE = 0.16$ ,  $p = .014$ ) and in the control task ( $t(67) = 2.24$ ,  $SE = 0.13$ ,  $p = .028$ ,  $t(67) = 1.56$ ,  $SE = 0.21$ ,  $p = .12$ , and  $t(67) = 1.07$ ,  $SE = 0.16$ ,  $p = .29$ ). Controlling for age did not qualitatively change these results. This activity in mostly posterior midline regions is consistent with previous research that identified the involvement of similar regions in prevention-focus processing (Johnson et al., 2006; Mitchell et al., 2009) and processing aging stereotype words for self-relevance (Colton et al., 2013). As can be seen in Fig. 3, while these results demonstrate the strongest effects in the episodic memory task, the effects were largely consistent across the three different tasks. To parallel the memory performance analysis, a MANOVA investigating the Group  $\times$  Education  $\times$  Retirement interaction was conducted on brain activity during the memory task in these three clusters. This analysis did not yield the same significant three-way interaction ( $p = .37$ ).

### 3.5. fMRI ROIs

We also analyzed brain activity in anatomical ROIs selected based on the two primary hypotheses guiding the stereotype threat literature: the executive-control interference hypothesis (bilateral amygdala, ACC, MFG, and left IFG/BA 47, see Krendl et al., 2008; Wraga et al., 2007) and a prevention focus under the regulatory fit hypothesis (MCC and PCC, see Johnson et al., 2006; Mitchell et al., 2009). We conducted two separate 2 (Group: stereotype, control)  $\times$  3 (Task: episodic encoding, 2-back, countdown) MANOVAs targeting each set of ROIs. The MANOVA on ROIs linked to the executive-control interference hypothesis did not yield a significant multivariate effect of Group or a Group  $\times$  Task interaction ( $p_s > 0.45$ ), and activity in these ROIs was quite similar across in the two groups (see Fig. S2). In contrast, the MANOVA on ROIs linked to the regulatory fit hypothesis yielded a significant effect of Group (Pillai's Trace = 0.11,  $F(4, 198) = 5.89$ ,  $p < .001$ ) but no interaction ( $p = .99$ ). To test which ROIs contributed to this overall effect, individual ANOVAs were conducted on each ROI, revealing significant group effects in right PCC ( $F(1201) = 13.52$ ,  $p < .001$ ), and left PCC ( $F(1201) = 10.18$ ,  $p < .001$ ), as shown in Fig. 4, but not the right MCC ( $p = .28$ ) nor the left MCC ( $p = .09$ , see Fig. S2). These PCC effects indicated greater brain activity in the stereotype group than the control group and were numerically strongest for brain activity in the episodic memory encoding task. These PCC patterns are quite similar to those found in the whole-brain analysis. Moreover, as seen in Fig. 4, participants in the control group showed the typical de-activation pattern in the PCC during the task blocks, consistent with a primary focus on the tasks as opposed to more inward, self-referential thoughts (e.g., Mak et al., 2017). In contrast, stereotype participants showed minimal de-activation, suggesting they remained self-focused throughout the tasks.

### 3.6. Brain-behavior relationships

To investigate brain-behavior relationships we used a multivariate method (PLS-R) that allowed the simultaneous assessment of multiple brain regions and multiple cognitive variables. In this PLS-R analysis, the significant clusters from the whole brain analysis were entered into one matrix and the cognitive measures (see Table 2) were entered as a second matrix except for response times to false alarms in the working memory task due to missing

values. For these analyses we collapsed across group status, because even though stereotype activation may have shifted participants' focus, participants in either group might have had prevention focus reactions to the experimental procedures. Two independent patterns explained much of the covariance in the data, explaining 86.05% and 10.94%, respectively (Fig. 5). The first latent variable was significant,  $r(67) = 0.36, p = .002$ , and indicated that greater brain activity in the right precentral sulcus and PCC (regardless of task) was associated with greater accuracy and lower false alarms across the episodic and working memory tasks. The second latent variable also was significant,  $r(67) = 0.28, p = .019$ , and indicated that greater brain activity in right precuneus (regardless of task) was associated with slower response times and lower false alarms to related lures (after correcting for responding to unrelated lures) in the episodic memory task. Together, the PLS-R results suggest that the greater activation found in brain regions that also were elevated in the Stereotype Group were associated with slower response times and greater accuracy across all participants, suggesting a more conservative, prevention-focus approach that is consistent with the regulatory fit hypothesis.

#### 4. Discussion

Here we report the first fMRI experiment to identify brain activity associated with stereotype activation while older adults took cognitive tasks. Two prominent hypotheses have been proposed to explain alterations in memory due to stereotype activation in older adults. The executive-control interference hypothesis emphasizes negative emotional reactions (performance anxiety) that compete with executive control resources, and this hypothesis has received support in the younger adult behavioral and neuroimaging literature on stereotype threat. The regulatory fit hypothesis instead emphasizes a strategy shift toward an error-prevention focus, and this hypothesis has received some support in the older adult behavioral literature on stereotype threat. The behavioral and neuroimaging findings reported here did not support the predictions of the executive-control interference hypothesis, suggesting that stereotype threat works differently in older adults compared to younger adults. Our behavioral and neuroimaging findings instead were more consistent with the regulatory fit hypothesis, although as we discuss below, additional work is now needed to replicate these MRI results and further understand their implications for cognitive processing during stereotype threat.

Several findings argue against the executive-control interference hypothesis, at least with respect to stereotype threat effects in older adults under the conditions used in our study. Previous studies have linked stereotype threat activation in younger adults to the anterior cingulate cortex and other regions that have been implicated in emotion regulation (Krendl et al., 2008; Wraga et al., 2007). We did not find group differences in these regions in either the whole-brain analysis or the targeted ROI analyses. The executive-control interference hypothesis also argues that working memory resources are needed to suppress negative emotions while taking cognitive tasks (Beilock et al., 2007; Schmader et al., 2008). Regions in prefrontal cortex, including ventrolateral PFC, have been implicated in cognitive control resources that might serve such a role. In the current study, the stereotype group did not recruit these brain regions to a greater extent than the control group. Furthermore, older adults in these two groups did not differ in a physiological assessment of parasympathetic

cardiac stress reactivity (HF HRV) or in reported anxiety, as would be expected from this hypothesis. Rather, both groups showed comparable increases in anxiety and HF HRV, which may reflect anticipatory responses to the cognitive tests and fMRI portion of the experiment. Finally, there were no differences between the groups in perceived threat from the experimental context itself. Because this questionnaire was given after the stereotype activation manipulation, the executive-control interference hypothesis would have predicted a difference on this measure.

In contrast to the executive-control interference hypothesis, several of our neuroimaging and behavioral findings were consistent with the regulatory fit hypothesis. Older adults in the stereotype group had greater activity in parietal midline regions than those in the control group. Greater activity in parietal midline regions, such as the PCC, has been associated with the activation of a prevention focus in both younger and older adults (Johnson et al., 2006; Mitchell et al., 2009), consistent with the regulatory fit hypothesis. Interestingly, EEG evidence shows that greater phase-locking associated with posterior midline regions (precuneus, PCC) and other regions in the default mode network (e.g., lateral parietal cortex) can help minimize stereotype threat effects on error monitoring and self-doubt in younger adults (Forbes et al., 2015). This finding suggests that activity in these posterior regions might signal a prevention focus that can help buffer against threat effects. In the current study, the group difference in PCC activity was observed across all the cognitive tasks, as would be expected if stereotype threat activated an error-prevention mindset throughout the experimental session. Additional support for the regulatory fit hypothesis stems from the behavioral findings. A shift to error prevention would arguably be most pronounced in false alarm rates. Although we did not find overall group differences in behavior, we did find that stereotype threat was most likely to impact episodic memory in older adults who are retired and more educated (cf. Smith et al., 2017). Moreover, this interaction was evident in the reduction in false alarms to new items during the episodic memory task. The brain-behavior correlations also showed that, pooling across our two experimental groups, those older adults showing the largest differences in brain activity in PCC also had the highest hit rates, lowest false alarm rates, and responded more slowly during the episodic memory task. Each of these patterns is consistent with a conservative and error-avoidant strategy.

One important consideration of the current fMRI findings is their generalizability across the different cognitive tasks. The whole-brain analysis yielded a group  $\times$  task interaction whereas the ROI analysis yielded a main effect of group across tasks, thus appearing to conflict with one another. However, inspecting the two patterns more closely reveals a quite similar pattern: the stereotype group effects were numerically strongest in the episodic memory encoding task, slightly less strong in the working memory task, and weakest in the countdown control task. There also was evidence that the stereotype group activated PCC more than the control group during the episodic retrieval task (see Supplementary Material). The brain-behavior correlations from the PLS-R analysis also yielded a graded pattern such that associations between PCC activity and performance were strongest for episodic memory task and weaker (but in the same direction) on the working memory task. It is unclear why stereotype effects on behavior and brain activity tended to be larger for the episodic memory encoding task than working memory task, but prior behavioral studies also have failed to find effects of stereotype activation on working memory performance in older adults (Hess

et al., 2009; Wong and Gallo, 2019). One possible explanation for the current pattern is that our working memory task was more tightly constrained than our episodic memory encoding task, in which participants were asked to memorize the words and but were not given a specific strategy or encoding judgment. Unconstrained tasks might be more sensitive to changes in motivation and strategy as proposed by the regulatory fit hypothesis, although additional work is needed on this point. The more definitive conclusion from the current dataset is that, although the effects of stereotype activation tended to be strongest on the episodic memory task, the primary effects on brain activity appeared to generalize across the tasks, suggesting a global shift in participants' overall approach to the tasks.

Related to this last point, it is important to realize that although activity in PCC during encoding might reflect a shift towards an error-prevention focus, this pattern does not imply that stereotype threat's effect on performance derives from its impact at encoding. In principle, explicitly activating stereotypes prior to the experimental tasks could have activated threat for the duration of the experimental session, so that alterations in either encoding or retrieval processes (or both) could have consequences for episodic memory performance. For example, a prevention focus might encourage participants to monitor memory more carefully for missing items from the studied categories, so as not to false alarm to these items, and these processes could have occurred while the categorized items were presented during encoding or retrieval. Depending on how effective older adults were at this kind of monitoring, it also might have caused them to be more sensitive to category membership when making memory judgments at test, thereby impacting false alarms to unrelated lures as well.

#### 4.1. Limitations and future directions

Several limitations in the current study highlight the need for additional research in this area. First, although the individual differences effects we observed in stereotype effects on episodic memory performance were motivated by prior behavioral work (Smith et al., 2017), this prior work itself has been mixed and our results do not resolve prior discrepancies. Additional behavioral work is needed to understand when and how explicit stereotype activation can impact cognitive performance in older adults. Second, the posterior midline regions that differed between the stereotype and control groups (e.g., PCC) have been implicated in a variety of functions other than prevention-focus (e.g., inward attention more generally), so that activity in PCC alone cannot be taken as the adoption of such a focus (see Poldrack, 2006, for the well-known problem of reverse inference in fMRI). Although our behavioral data in conjunction with the neuroimaging data were more consistent with the regulatory fit hypothesis than the executive-control interference hypothesis, few prior fMRI studies have investigated the neural correlates of adopting a prevention focus (e.g., Johnson et al., 2006; Mitchell et al., 2009). Thus, even though activity in this region is consistent with the regulatory fit hypothesis when considered with the behavioral evidence, additional work is needed to understand the functional role of this region during stereotype activation in older adults. A third limitation is that, although we randomized participants into two groups and they did not differ on neuropsychological assessments (MoCA or Shipley Scales), the stereotype group was somewhat older and had lower false recognition in the baseline session (prior to any manipulation). While analyses controlling for age yielded similar results as our



primary analyses, the possibility of unintended group differences is always a possibility in a between-subjects design.

A final caveat is that our study does not rule out the possibility that older adults may experience anxiety and executive-control interference from activation of aging stereotypes in other contexts. For example, in our study there was evidence that the MRI environment may have been somewhat stressful or anxiety-provoking to older adults in both conditions. This anxiety may have minimized or overshadowed our ability to find group differences in emotional regulation regions associated with stereotype activation predicted by the executive-control interference hypothesis. Although our participants had a pre-MRI baseline session, which should have reduced overall test-taking anxiety during the MRI session, we cannot rule out this possibility. We are unaware of any work aimed at comparing the behavioral effects of stereotype activation in the MRI environment versus other contexts, although one study has shown that memory performance, generally, can be impaired in the MRI compared with a laboratory environment (Gutchess and Park, 2006). Future work could investigate potential interactions between such context effects and stereotype activation effects.

In conclusion, activating aging stereotypes increased brain activity in parietal midline regions associated with the processing of aging stereotypes, the self, and an error-prevention focus. While activity in these brain regions was consistent with regulatory fit hypothesis, we did not consistently find patterns of brain activity that were predicted by the executive-control interference hypothesis. These results add new constraints on current hypotheses of stereotype threat effects as they apply to older adults. Specifically, growing evidence in the behavioral literature suggests that the mechanisms of stereotype threat in older adults may be different from those in younger adults. Our results resonate with this emerging picture, demonstrating that the impact of stereotype threat on brain activity in older adults' also is different from what one might expect from the relevant neuroimaging literature in younger adults. Aging stereotype threat seems to operate differently than other kinds of stereotype threat, at both the behavioral and brain levels of analysis.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments and funding sources

Research reported in this publication was supported by the [National Institute on Aging](#) of the National Institutes of Health under Award Number R21AG049931. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Data/code availability

Neuroimaging and behavioral data used in this study will be freely available on the Open Science Framework upon publication. The software and models that we used to generate the results are stated in appropriate locations in the Methods section.

## Appendix.:

### Stereotype and Control Passages

#### Stereotype Passage:

Thank you for participating in our study. Today, you will take memory tasks like the ones that you previously took in the lab. We will measure your performance on these tasks, and we also will take pictures of your brain's structures and your brain's activity during the tasks. Scientific studies have shown that these procedures can detect memory decline associated with normal aging, and they also can predict the development of Alzheimer's disease later in life.

A main goal of our study is to increase the effectiveness of these measures at detecting memory decline associated with normal aging and Alzheimer's disease. To achieve this goal, we will compare data from young college students to older adults in your age group. In addition, we will compare data between older adults whose performance puts them at different risk levels for the development of AD later in life. Based on your data from Day 1, our recruiters have determined that you are eligible for this study. Given these goals, it is important that you understand the memory task instructions that I will now give you.

#### Control Passage:

Thank you for participating in our study. Today, you will take psychological tasks like the ones that you previously took in the lab. We will measure your performance on these tasks, and we also will take pictures of your brain's structures and your brain's activity during the tasks. Scientific studies have shown that these procedures can help us understand how the brain supports different psychological processes.

A main goal of this study is to evaluate brain activity associated with individual differences in psychological processes that make each of us unique. To achieve this goal, we will compare data across a large sample of research participants with varying backgrounds and characteristics. Based on your data from Day 1, our recruiters have determined that you are eligible for this study. Given these goals, it is important that you understand the psychology task instructions that I will now give you.

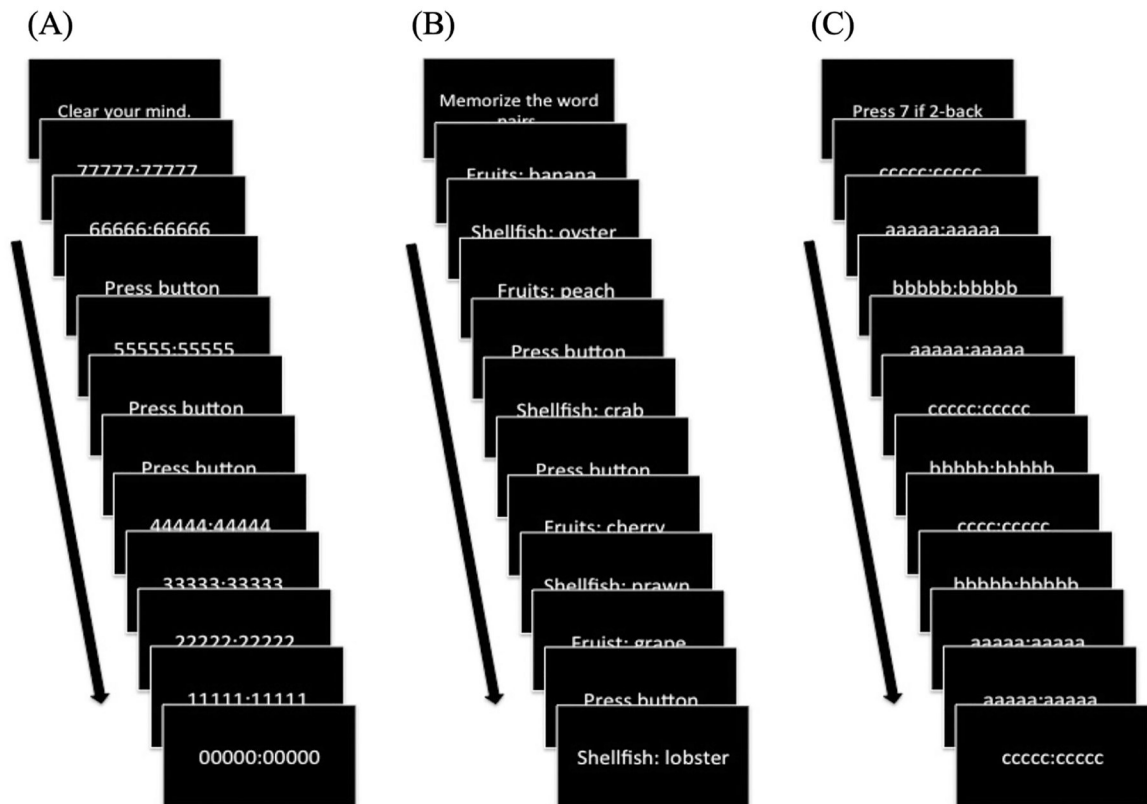
### References

- Armstrong B, Gallant SN, Li L, Patel K, Wong BI, 2017. Stereotype threat effects on older adults' episodic and working memory: a meta-analysis. *Gerontologist* 57 (suppl\_2), S0193–S205.
- Barber SJ, Mather M, 2013a. Stereotype threat can both enhance and impair older adults' memory. *Psychol. Sci* 24 (12), 2522–2529. [PubMed: 24150969]
- Barber SJ, Mather M, 2013b. Stereotype threat can reduce older adults' memory errors. *Q. J. Exp. Psychol* 66 (10), 1888–1895.
- Barber SJ, Mather M, 2014. Stereotype threat in older adults: when and why does it occur and who is most affected?. Stereotype threat in older adults: when and why does it occur and who is most affected? In: Verhaeghen P, Hertzog C (Eds.) *The Oxford Handbook of Emotion, Social Cognition, and Problem Solving in Adulthood*. Oxford University Press, pp. 302–319.
- Barber SJ, 2017. An examination of age-based stereotype threat about cognitive decline: implications for stereotype-threat research and theory development. *Perspect. Psycholog. Sci* 12 (1), 62–90.

- Beaton D, Fatt CRC, Abdi H, 2014. An ExPosition of multivariate analysis with the singular value decomposition in R. *Comput. Stat. Data Anal* 72, 176–189.
- Beilock SL, Rydell RJ, McConnell AR, 2007. Stereotype threat and working memory: mechanisms, alleviation, and spillover. *J. Exper. Psychol* 136 (2), 256.
- Berntson GG, Quigley KS, Norman GJ, Lozano DL, 2017. Cardiovascular psychophysiology. In: Cacioppo JT, Tassinary LG, Berntson GG (Eds.), *Handbook of Psychophysiology*. Cambridge University Press, pp. 183–216.
- Colton G, Leshikar ED, Gutchess AH, 2013. Age differences in neural response to stereotype threat and resiliency for self-referenced information. *Front. Hum. Neurosci* 7, 537. [PubMed: 24046739]
- Colzato LS, Steenbergen L, 2017. High vagally mediated resting-state heart rate variability is associated with superior action cascading. *Neuropsychologia* 106, 1–6. [PubMed: 28866318]
- Derks B, Inzlicht M, Kang S, 2008. The neuroscience of stigma and stereotype threat. *Group Process. Intergroup Relation* 11, 163–181.
- Du YP, Dalwani M, Wylie K, Claus E, Tregellas JR, 2007. Reducing susceptibility artifacts in fMRI using volume-selective z-shim compensation. *Magnet. Reson. Med* 57 (2), 396–404.
- Forbes CE, Leitner JB, 2014. Stereotype threat engenders neural attentional bias towards negative feedback to undermine performance. *Biol. Psychol* 102, 98–107. [PubMed: 25063472]
- Forbes CE, Leitner JB, Duran-Jordan K, Magerman AB, Schmader T, Allen JJB, 2015. Spontaneous default mode network phase-locking moderates performance perceptions under stereotype threat. *Soc. Cogniti. Affect. Neurosci* 10, 994–1002.
- Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RSJ, 1995. Statistical parametric maps: confidence intervals on p-values. *Hum. Brain Mapp* 2, 189–210.
- Gallo DA, 2004. Using recall to reduce false recognition: diagnostic and disqualifying monitoring. *J. Exper. Psychol* 30, 120–128.
- Gudjonsson GH, 1989. Compliance in an interrogative situation: a new scale. *Pers. Individ. Dif* 10 (5), 535–540.
- Gutchess AH, Park DC, 2006. fMRI environment can impair memory performance in young and elderly adults. *Brain Res* 1099 (1), 133–140. [PubMed: 16765326]
- Hasher L, Zacks RT, Bower GH (Ed.), 1988. Working memory, comprehension, and aging: a review and a new view. *The Psychology of Learning and Motivation* 193–225 Vol. 22(1988).
- Haslam C, Morton TA, Haslam SA, 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. *Psychol. Aging* 27 (3), 778. [PubMed: 22468850]
- Hess TM, Auman C, Colcombe SJ, Rahhal TA, 2003. The impact of stereotype threat on age differences in memory performance. *J. Gerontol. Ser. B* 58 (1), P3–P11.
- Hess TM, Hinson JT, Hodges EA, 2009. Moderators of and mechanisms underlying stereotype threat effects on older adults’ memory performance. *Exp. Aging Res* 35 (2), 153–177. [PubMed: 19280445]
- Higgins ET, 1998. Promotion and prevention: regulatory focus as a motivational principle. *Adv. Exp. Soc. Psychol* 30, 1–46.
- Johnson MK, Raye CL, Mitchell KJ, Touryan SR, Greene EJ, Nolen-Hoeksema S, 2006. Dissociating medial frontal and posterior cingulate activity during self-reflection. *Soc. Cogn. Affect. Neurosci* 1 (1), 56–64. [PubMed: 18574518]
- Kang SK, Chasteen AL, 2009. The moderating role of age-group identification and perceived threat on stereotype threat among older adults. *Int. J. Aging Hum. Dev* 69 (3), 201–220. [PubMed: 20041566]
- Krendl AC, Richeson JA, Kelley WM, Heatherton TF, 2008. The negative consequences of threat: a functional magnetic resonance imaging investigation of the neural mechanisms underlying women’s underperformance in math. *Psychol. Sci* 19 (2), 168–175. [PubMed: 18271865]
- Kelley WM, Miezin FM, McDermott KB, Buckner RL, Raichle ME, Cohen NJ, Ollinger JM, Akbudak E, Conturo TE, Snyder AZ, Petersen SE, 1998. Hemispheric specialization in human dorsal frontal cortex and medial temporal lobe for verbal and nonverbal memory encoding. *Neuron* 20 (5), 927–936. [PubMed: 9620697]

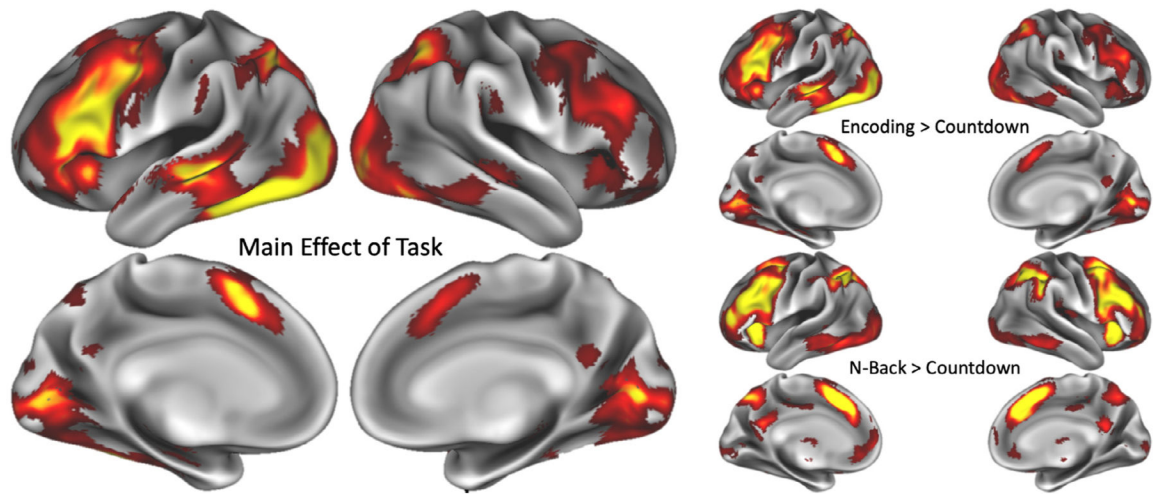
- Lachman ME, Bandura M, Weaver SL, Elliott E, 1995. Assessing memory control beliefs: the memory controllability inventory. *Aging Neuropsychol. Cognit* 2 (1), 67–84.
- Lamont RA, Swift HJ, Abrams D, 2015. A review and meta-analysis of age-based stereotype threat: negative stereotypes, not facts, do the damage. *Psychol. Aging* 30 (1), 180. [PubMed: 25621742]
- Mak LE, Minuzzi L, MacQueen G, Hall G, Kennedy SH, Milev R, 2017. The default mode network in healthy individuals: a systematic review and meta-analysis. *Brain Connect* 7 (1), 25–33. [PubMed: 27917679]
- Mangels JA, Good C, Whiteman RC, Maniscalco B, Dweck CS, 2012. Emotion blocks the path to learning under stereotype threat. *Soc. Cogn. Affect. Neurosci* 7, 230–241. [PubMed: 21252312]
- Mazerolle M, Régner I, Morisset P, Rigalleau F, Huguet P, 2012. Stereotype threat strengthens automatic recall and undermines controlled processes in older adults. *Psychol. Sci* 23 (7), 723–727. [PubMed: 22609539]
- Meisner BA, 2012. A meta-analysis of positive and negative age stereotype priming effects on behavior among older adults. *J. Gerontol. Ser. B* 67 (1), 13–17.
- Meyer TJ, Miller ML, Metzger RL, Borkovec TD, 1990. Development and validation of the Penn State Worry Questionnaire. *Behav. Res. Therapy* 28, 487–495.
- Mitchell KJ, Raye CL, Ebner NC, Tubridy SM, Frankel H, Johnson MK, 2009. Age-group differences in medial cortex activity associated with thinking about self-relevant agendas. *Psychol. Aging* 24, 438–449. [PubMed: 19485660]
- Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H, 2005. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc* 53 (4), 695–699. [PubMed: 15817019]
- Nee DE, Brown JW, Askren MK, Berman MG, Demiralp E, Krawitz A, Jonides J, 2013. A meta-analysis of executive components of working memory. *Cereb. Cortex* 23 (2), 264–282. [PubMed: 22314046]
- Oldfield RC, 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9 (1), 97–113. [PubMed: 5146491]
- Owen AM, McMillan KM, Laird AR, Bullmore E, 2005. N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Hum. Brain Mapp* 25 (1), 46–59. [PubMed: 15846822]
- Pennington CR, Heim D, Levy AR, Larkin DT, 2016. Twenty years of stereotype threat research: a review of psychological mediators. *PLoS One* 11 (1), e0146487. [PubMed: 26752551]
- Poldrack RA, 2006. Can cognitive processes be inferred from neuroimaging data? *Trend. Cognit. Sci* 10, 59–63.
- Porges SW, 2001. The polyvagal theory: phylogenetic substrates of a social nervous system. *Int. J. Psychophysiol* 42 (2), 123–146. [PubMed: 11587772]
- Schmader T, Johns M, Forbes C, 2008. An integrated process model of stereotype threat effects on performance. *Psychol. Rev* 115 (2), 336. [PubMed: 18426293]
- Seibt B, Förster J, 2004. Stereotype threat and performance: how self-stereotypes influence processing by inducing regulatory foci. *J. Pers. Soc. Psychol* 87 (1), 38. [PubMed: 15250791]
- Shibley WC, Gruber CP, Martin TA, Klein AM, 2009. Shipley-2. Western Psychological Services, Los Angeles, CA.
- Smith AM, Gallo DA, Barber SJ, Maddox KB, Thomas AK, 2017. Stereotypes, warnings, and identity-related variables influence older adults' susceptibility to associative false memory errors. *Gerontologist* 57 (suppl\_2), S206–S215. [PubMed: 28854608]
- Spaniol J, Davidson PS, Kim AS, Han H, Moscovitch M, Grady CL, 2009. Event-related fMRI studies of episodic encoding and retrieval: meta-analyses using activation likelihood estimation. *Neuropsychologia* 47 (8–9), 1765–1779. [PubMed: 19428409]
- Spencer SJ, Logel C, Davies PG, 2016. Stereotype threat. *Annu. Rev. Psychol* 67, 415–437. [PubMed: 26361054]
- Spielberger CD, 2010. State-Trait Anxiety Inventory. John Wiley & Sons, Inc.
- Thalbourne M, Delin P, 1993. A new instrument for measuring the sheep-goat variable: its psychometric properties and factor structure. *J. Soc. Psych. Res* 59, 172–186.

- Thayer JF, Hansen AL, Saus-Rose E, Johnsen BH, 2009. Heart rate variability, prefrontal neural function, and cognitive performance: the neurovisceral integration perspective on self-regulation, adaptation, and health. *Ann. Behav. Med* 37 (2), 141–153. [PubMed: 19424767]
- Thomas AK, Dubois SJ, 2011. Reducing the burden of stereotype threat eliminates age differences in memory distortion. *Psychol. Sci* 22 (12), 1515–1517. [PubMed: 22030349]
- Thomas AK, Smith AM, Mazerolle M, 2020. The unexpected relationship between retrieval demands and memory performance when older adults are faced with age-related stereotypes. *J. Gerontol. Ser. B* 75 (2), 241–250.
- Trapnell PD, Campbell JD, 1999. Private self-consciousness and the five-factor model of personality: distinguishing rumination from reflection. *J. Pers. Soc. Psychol* 76 (2), 284. [PubMed: 10074710]
- Williams DP, Joseph N, Hill LK, Sollers III JS, Vasey MW, Way BM, Koenig J, Thayer JF, 2019. Stereotype threat, trait perseveration, and vagal activity: evidence for mechanisms underpinning health disparities in Black Americans. *Ethn. Health* 24, 909–926. [PubMed: 28922935]
- Wong JT, Gallo DA, 2016. Stereotype threat reduces false recognition when older adults are forewarned. *Memory* 24 (5), 650–658. [PubMed: 26218527]
- Wong JT, Gallo DA, 2019. Activating aging stereotypes increases source recollection confusions in older adults: effect at encoding but not retrieval. *J. Gerontol. Ser. B* 74 (4), 633–641.
- Wraga M, Helt M, Jacobs E, Sullivan K, 2007. Neural basis of stereotype-induced shifts in women's mental rotation performance. *Soc. Cogn. Affect. Neurosci* 2 (1), 12–19. [PubMed: 18985116]

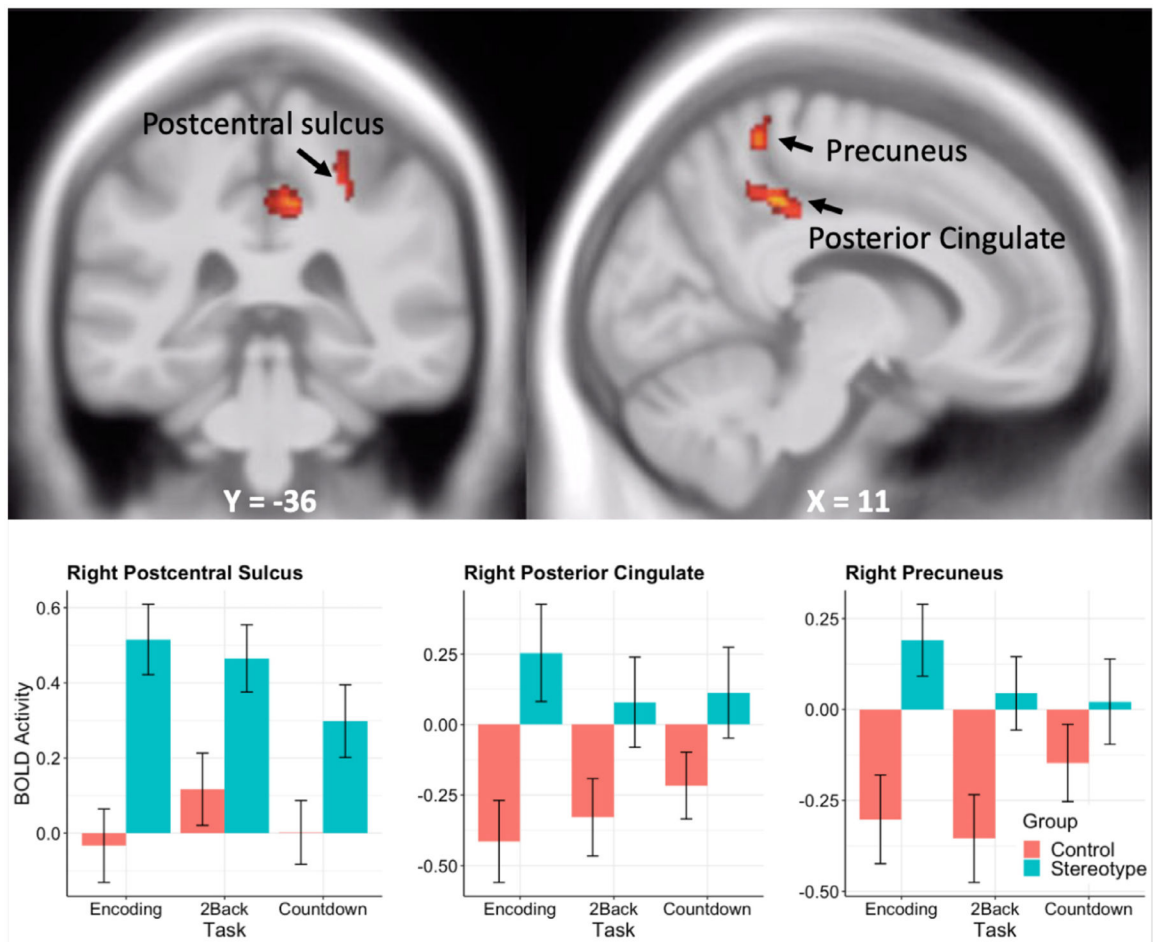


**Fig. 1.**

Schematic of Task Procedure. Participants were given the cognitive tasks in the baseline session, as well as in the subsequent session (with new sets of items) in which stereotype activation was manipulated just prior to taking the tasks during MRI. The tasks were presented across multiple, interleaved blocks: (A) a control task where participants read a numerical countdown, (B) an episodic memory encoding task for words, and (C) a working memory task on letters (the 2-back task). The blocked tasks were matched in number of trials, number of button presses, and visual complexity. Following the blocked tasks, participants took a recognition memory test for the encoded words using an event-related fMRI design.

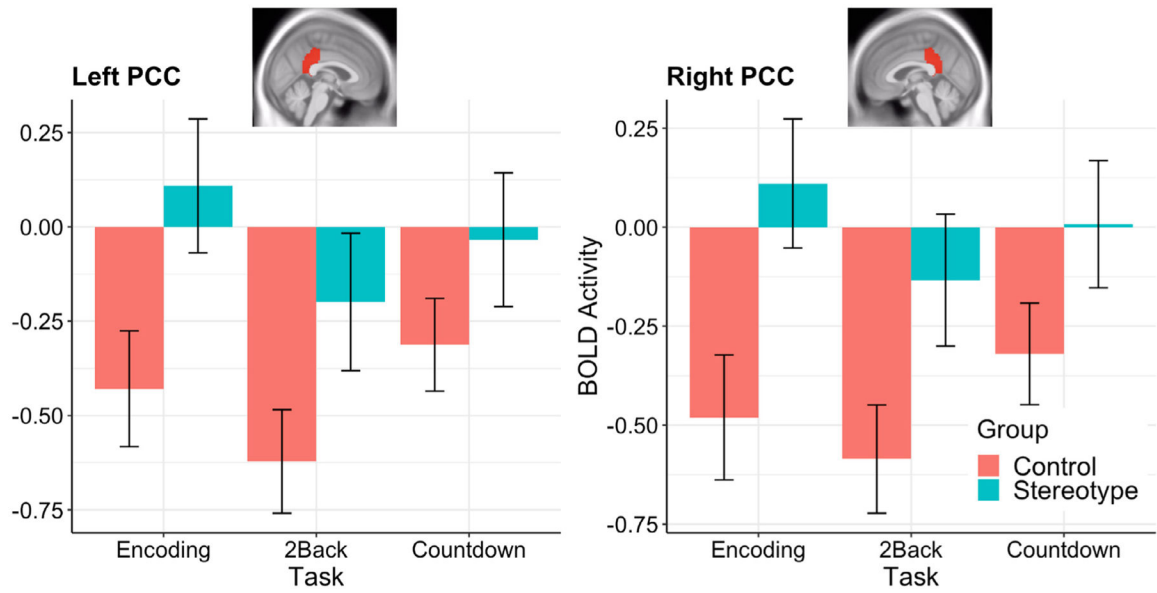


**Fig. 2.** Whole Brain Results Illustrating Task Block Effects. The left panel shows the main effect of task in the mixed ANOVA. The right panel shows separate effects for the episodic memory encoding task (top) and working memory task (bottom). Hot colors represent greater brain activity for the task. No deactivations were found.

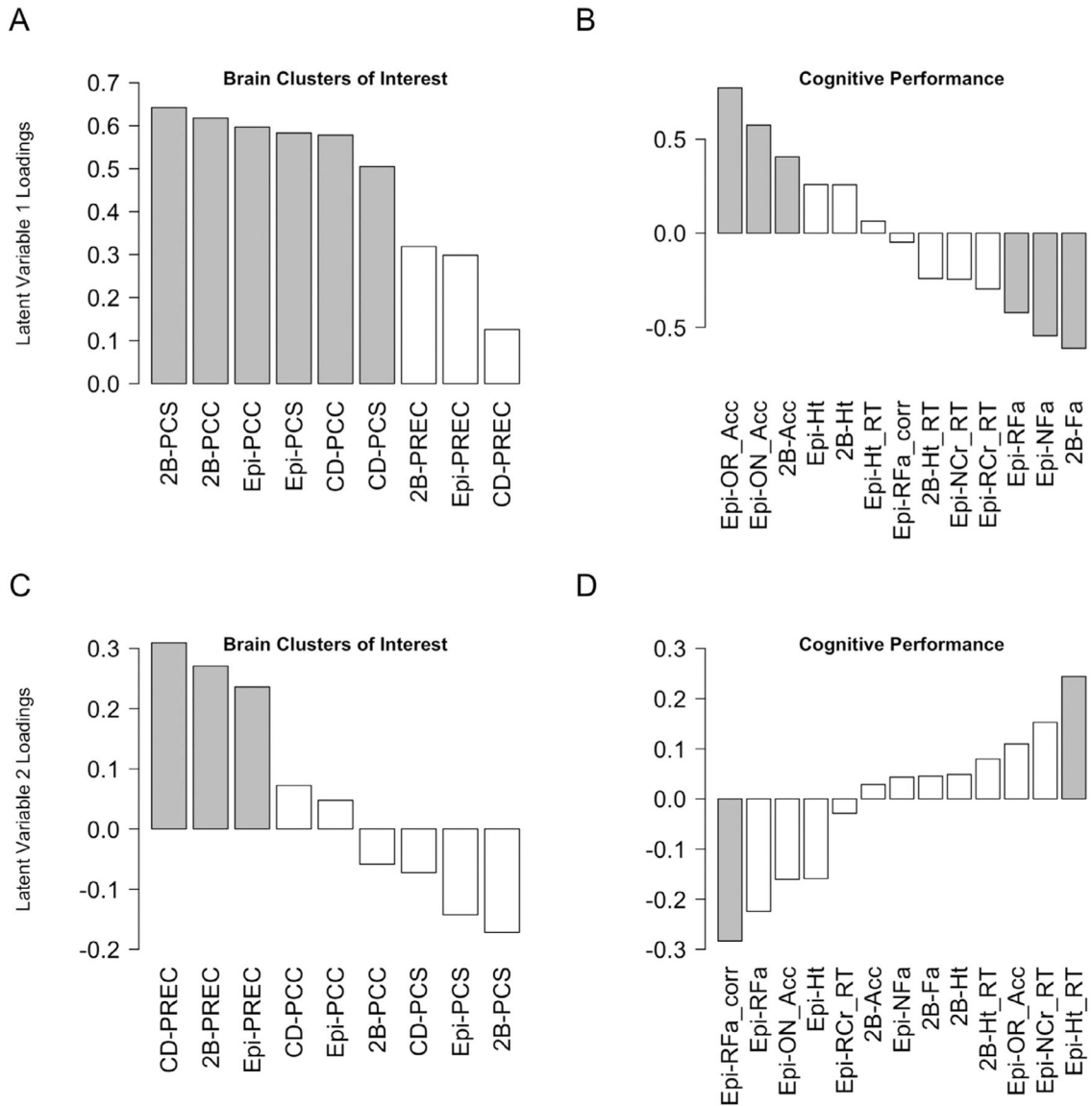


**Fig. 3.** Whole Brain Results Illustrating Stereotype Group  $\times$  Blocked Task Interaction. The top panel represents the three clusters of brain activity showing the interaction. The bottom panel shows bar plots with parameter estimates for each of the three cognitive tasks for the Control Group (red) and Stereotype Group (blue). Error bars represent the standard error.





**Fig. 4.** Anatomical ROI Results Illustrating the Main Effect of Group on the Task Blocks. Significant group effects were found in both the left and right PCC, but not in other ROIs (Fig S2). Bar plots are shown with parameter estimates for each of the three cognitive tasks for the Control Group (red) and Stereotype Group (blue). Error bars represent the standard error. PCC = posterior cingulate cortex; Encoding = Episodic memory task during encoding.



**Fig. 5.** Barplots of the Partial Least Squares Regression Analyses of Brain-Behavior Relationships. Panel A shows brain activity most expressed in the first latent variable. Panel B shows the cognitive measures most expressed in the first latent variable. Panel C shows brain activity most expressed in the second latent variable. Panel D shows the cognitive measures most expressed in the second latent variable. Gray bars indicate the greatest factor loadings for each latent variable. PCC = posterior cingulate cortex; PCS = postcentral sulcus; PREC = precuneus; Epi = episodic memory task (encoding activity and retrieval behavior); 2B = 2-Back task; CD = countdown task; Acc = accuracy; ON = old/new; OR = old/related; Ht = hit rate; RFa = related lure false alarm rate; NFa = new (unrelated) lure false alarm rate; RFa\_corr = corrected false alarm rate; RCr = related lure correct rejection; NCr = new (unrelated) lure correct rejection; RT = response time.

**Table 1**  
Sociodemographic Variables, Neuropsychological Tests, Psychological Questionnaires, and Physiological Measures for Each Group.

	Stereotype ( <i>n</i> = 33) M (SD) or N (%)	Control ( <i>n</i> = 36) M (SD) or N (%)	<i>T</i> - or $\chi^2$ <i>p</i> -value
Age	64.61 (2.87)	62.72 (2.72)	.007
Sex (female)	22 (67%)	20 (56%)	.49
Years of Education	14.88 (2.48)	15.08 (2.57)	.74
Ethnoracial Category (non-Hispanic White)	10 (30%)	14 (39%)	0.62
Retired	20 (61%)	19 (53%)	.68
Income Scale	4.59 (3.58)	4.86 (3.63)	.76
MoCA	25.53 (2.40)	25.19 (3.30)	.63
Shipley Vocabulary	32.55 (5.06)	30.41 (7.57)	.17
Shipley Logic	22.56 (9.18)	23.17 (9.59)	.79
Rumination Tendencies	30.06 (7.88)	34.83 (9.20)	.02
Worry Tendencies	34.84 (10.87)	38.94 (11.40)	.13
Perceived Stereotype Threat (Total)	11.19 (4.20)	13.36 (3.87)	.03
Perceived Stereotype Threat (Experiment)	1.47 (0.95)	1.67 (0.93)	.39
Trait Anxiety	34.72 (8.18)	36.33 (8.85)	.44
Memory Controllability Inventory (Total)	86.06 (5.35)	83.78 (7.41)	.15
State Anxiety			
Baseline	27.09 6.00	26.89 6.89	.90
Post	36.38 10.12	35.44 7.87	.68
HF Heart Rate Variability			
Baseline	4.12 (1.28)	4.22 (1.63)	.80
Passage	4.19 (1.23)	4.22 (1.29)	.91
Post	4.46 (1.25)	4.47 (1.44)	.98

Notes. The six demographic variables, MoCA, Shipley Scales, and questionnaires on Rumination Tendencies, Worry Tendencies, and Trait Anxiety were administered at end of the first session (no MRI). Memory Controllability Inventory and Perceived Stereotype Threat questionnaires were assessed at the end of the second session (after MRI). State anxiety and heart rate specific were assessed immediately before the stereotype passage manipulation (pre) and after the MRI session (post), with heart rate also measured while reading the passage (see text).

Table 2

Cognitive Performance for Each Group in Each Testing Session.

	Non-MRI Session (no manipulation)		MRI Session (with manipulation)		<i>T</i> -test <i>p</i> -value
	Stereotype ( <i>n</i> = 33)	Control ( <i>n</i> = 36)	Stereotype ( <i>n</i> = 33)	Control ( <i>n</i> = 36)	
<b>Episodic Memory</b>					
Hits (%)	.68 (0.17)	.68 (0.17)	.71 (0.17)	.71 (0.19)	.91
Related Lure FA (%)	.40 (0.18)	.46 (0.19)	.42 (0.20)	.51 (0.22)	.09
New FA (%)	.11 (0.11)	.16 (0.13)	.14 (0.12)	.19 (0.16)	.16
Old/New Accuracy (%)	.58 (0.18)	.51 (0.20)	.57 (0.20)	.52 (0.23)	.40
Old/Related Accuracy (%)	.28 (0.14)	.22 (0.14)	.29 (0.16)	.20 (0.18)	.05
Corrected FA (%)	.29 (0.16)	.29 (0.17)	.28 (0.19)	.32 (0.19)	.45
Hits (RT)	1530.93 (256.58)	1562.50 (254.08)	1453.64 (181.94)	1443.96 (246.19)	.85
Related Lure CR (RT)	1736.49 (284.91)	1762.60 (268.78)	1668.61 (231.02)	1632.30 (268.25)	.55
New CR (RT)	1540.96 (283.63)	1545.64 (263.74)	1475.36 (195.65)	1428.88 (286.08)	.43
<b>Working memory</b>					
Hits (%)	.71 (0.17)	.71 (0.19)	.69 (0.22)	.69 (0.19)	.97
FA (%)	.09 (0.08)	.09 (0.08)	.09 (0.07)	.08 (0.07)	.57
Accuracy (%)	.63 (0.19)	.62 (0.21)	.60 (0.25)	.61 (0.21)	.85
Hits (RT)	793.3 (161.94)	799.38 (161.12)	825.10 (173.9)	844.81 (167.04)	.63
FA (RT) *	800.08 (239.9)	865.98 (248.82)	877.88 (264.23)	848.92 (259.23)	.67

Notes. Means are reported with standard deviations in parentheses. FA = False Alarms; RT = Response Times; CR = Correct Rejections.

\* 10 participants (5 per group) had no false alarms and so response times could not be calculated.

**Table 3**

Whole Brain fMRI Results Showing Group  $\times$  Task Interactions.

MINI Coordinates (x, y, z)	Region	BA	Z-Value	Cluster Size (voxels)
10, -40, 42	Right Posterior Cingulate Gyrus	31	3.78	156
12, -46, 64	Right Precuneus	5	3.77	129
24, -46, 46	Right Postcentral Sulcus	3	4.24	127