THE GERONTOLOGICAL SOCIETY OF AMERICA®

*Journals of Gerontology: Psychological Sciences*
cite as: *J Gerontol B Psychol Sci Soc Sci*, 2019, Vol. 74, No. 4, 633–641
doi:10.1093/geronb/gbx103
Advance Access publication March 16, 2018

OXFORD

Original Research Report

# Activating Aging Stereotypes Increases Source Recollection Confusions in Older Adults: Effect at Encoding but Not Retrieval

**Jessica T. Wong, PhD and David A. Gallo, PhD**

Department of Psychology, University of Chicago, Illinois.

Address correspondence to: David A. Gallo, PhD, Department of Psychology, University of Chicago, 5848 S University Ave, Chicago, IL 60637. E-mail: dgallo@uchicago.edu

Received: December 11, 2016; Editorial Decision Date: June 28, 2017

**Decision Editor:** Bob G. Knight, PhD

## Abstract

**Objectives:** Activating aging stereotypes can impair older adult performance on episodic memory tasks, an effect attributed to stereotype threat. Here, we report the first study comparing the effects of explicitly activating aging stereotypes at encoding versus retrieval on recollection accuracy in older adults.

**Method:** During the encoding phase, older adults made semantic judgments about words, and during the retrieval phase, they had to recollect these judgments. To manipulate stereotype activation, participants read about aging-related decline (stereotype condition) or an aging-neutral passage (control condition), either before encoding or after encoding but before retrieval. We also assessed stereotype effects on metacognitive beliefs and two secondary tasks (working memory, general knowledge) administered after the recollection task.

**Results:** Stereotype activation at encoding, but not retrieval, significantly increased recollection confusion scores compared to the control condition. Stereotype activation also increased self-reports of cognitive decline with aging, but it did not reliably impact task-related metacognitive assessments or accuracy on the secondary tasks.

**Discussion:** Explicitly activating aging stereotypes at encoding increases the likelihood of false recollection in older adults, potentially by diminishing encoding processes. Stereotype activation also influenced global metacognitive assessments, but this effect may be unrelated to the effect of stereotypes on recollection accuracy.

**Keywords:** Episodic memory, False memory, Metacognition, Semantic memory, Working memory

Aging is associated with memory impairments and changes in underlying brain systems (Grady, 2012), but research indicates that negative aging stereotypes also may contribute to these performance declines (e.g., Levy, Zonderman, Slade, & Ferrucci, 2012). Stereotype threat is the idea that activating negative beliefs about a stigmatized group causes group members to underperform (Steele & Aronson, 1995). Many experiments with cognitively normal older adults have found that activating negative aging stereotypes can impair cognitive performance (for reviews, see Barber, 2017; Lamont, Swift, & Abrams, 2015). Most research in this area has focused on episodic memory tasks, and has explicitly activated stereotypes by having older adults read passages about aging-related cognitive decline (or a control passage) just prior to the task. This procedure typically reduces recall or recognition memory for word lists (e.g., Hess, Auman, Colcombe, & Rahhal, 2003; Hess, Emery, & Queen, 2009a), suggesting a negative impact of stereotype threat on episodic memory in older adults. For consistency with the literature, we use the term "stereotype threat" to refer to the psychological reactions that people might experience when confronted with explicit stereotypes. Our use of the term "stereotype" is not meant to imply that changes in the aging brain do not drive aspects of aging-related cognitive decline and associated stereotypes, nor is our use of the term "threat" meant to imply that everyone feels threatened by the explicit activation of these stereotypes.

Several theories try to explain how stereotype activation impairs memory performance in older adults (see Barber & Mather, 2014; Popham & Hess, 2015b). One hypothesis is that the activation of negative stereotypes preoccupies older adults with stereotype-related worries, thereby reducing cognitive processes devoted to the task (e.g., Mazerolle, Régner, Morisset, Rigalleau, & Huguet, 2012). Another hypothesis is that activating stereotypes causes older adults to alter their processing strategies, even if processing resources remain intact. For example, stereotypes might decrease older adults' motivation to cognitively engage the encoding task, or they might motivate older adults to be more cautious and respond more conservatively under typical testing situations (see Barber & Mather, 2013a). The relative role of these different mechanisms is still a topic of debate, and multiple processes might contribute to stereotype effects on memory.

One way to gain theoretical insight into these underlying mechanisms is to disentangle the effect of stereotype activation on memory encoding and retrieval processes. The overwhelming majority of stereotype threat studies in older adults has activated stereotypes before the encoding phase of a memory task, so that either encoding or retrieval processes could have been impacted (e.g., Barber & Mather, 2013b; Hess et al., 2003; Hess et al., 2009a; Hess, Hinson, & Hodges, 2009b; Hess & Hinson, 2006; Kang & Chasteen, 2009; Mazerolle et al., 2012). With respect to encoding processes, stereotype activation could reduce the likelihood that older adults would self-initiate elaborative encoding processes that benefit memory (e.g., Craik & Tulving, 1975). With respect to retrieval processes, stereotype activation could reduce the effortful search for previously stored memories, as well as the use of postretrieval monitoring processes to regulate memory accuracy (e.g., Johnson & Raye, 1981).

To our knowledge, only three studies have explicitly activated aging stereotypes after the encoding phase but before the memory task, so that only retrieval could be affected. Thomas and Dubois (2011) found that activating stereotypes just before retrieval increased false recognition errors in the DRM task (Roediger & McDermott, 1995), whereas Wong and Gallo (2016) found the opposite result with a modified DRM procedure. In a follow-up study that compared the two DRM procedures, Smith, Gallo, Barber, Maddox, & Thomas (in press) obtained results that were more consistent with Thomas and Dubois (2011). These studies indicate that stereotype activation prior to retrieval can impact memory errors, potentially by diminishing retrieval monitoring accuracy. However, these effects have been inconsistent, potentially because it is difficult to control retrieval strategies in the DRM task (see Wong & Gallo, 2016). Moreover, these prior studies did not include a condition where stereotypes were activated prior to encoding, so the effect of activating stereotypes at encoding could not be evaluated.

In the current study, we assessed stereotype effects using a source recollection task that afforded better control of retrieval strategies. We explicitly activated stereotypes prior to the study phase (encoding-stereotype condition) or after the study phase but prior to the test phase (retrieval-stereotype condition). Whereas activating stereotypes prior to the study phase could affect either encoding or retrieval processes, activating stereotypes after the study phase could affect retrieval processes alone. If aging stereotypes impact encoding more than retrieval processes, activating stereotypes at encoding should have a larger effect than retrieval. By contrast, if aging stereotypes impact memory retrieval more than encoding processes, activating stereotypes at retrieval might have the larger effect.

Although this is the first study to directly compare the effects of explicitly activating aging stereotypes at encoding and retrieval, a study by Krendl, Ambady, & Kensinger (2015) used a subliminal procedure to activate aging stereotypes at encoding or retrieval. That study found that activating stereotypes increased false recognition in older adults, with larger effects when stereotypes were activated just prior to retrieval. These findings support the hypothesis that stereotype activation impacts retrieval more than encoding, although the subliminal activation of stereotypes may operate under different mechanisms than the explicit activation of stereotypes that are of interest in the present study.

In addition to investigating the effect of stereotype activation on different stages of episodic memory, a second goal of our study was to investigate the extent that stereotype activation impacts metacognitive expectations about ones own cognitive abilities. Inducing stereotype threat can reduce feelings of self-efficacy or performance expectations in older adults (e.g., Chasteen, Bhattacharyya, Horhota, Tam, & Hasher, 2005; Desrichard & Kopetz, 2005; Hess et al., 2009b), and a recent study by Bouazzaoui et al. (2016) found that the negative impact of explicit stereotype activation on word recall was mediated by its effect on older adults' ratings of memory self-efficacy. These findings suggest that stereotype threat might impact performance by altering older adults' metacognitive beliefs about their own cognitive abilities, which in turn could impact their approach to the task (e.g., their use of encoding or retrieval strategies). We tested this hypothesis by comparing the effects of stereotype activation on self-rated metacognitive ability in three memory domains (episodic memory, working memory, and semantic memory or general knowledge). We also administered working memory and general knowledge tasks to assess stereotype effects on actual performance in these domains, although these tasks always followed the episodic memory task, so that the stereotype manipulation may have been weaker for these secondary tasks relative to the episodic memory task.

## Methods

### Participants

One hundred and twenty-six older adults participated (age 65–90 years; $M = 75.25$, $SD = 6.36$). All older adults lived independently in the Chicago area and were screened for cognitive or affective problems impairing daily functioning, using the Mini-Mental Status Exam (Folstein, Folstein, & McHugh, 1975) and the Geriatric Depression Scale (GDS; Brink et al., 1982). There were three between-subjects conditions, with 42 participants in each: stereotypes activated before the episodic memory task encoding phase (encoding-stereotype condition), stereotypes activated after encoding but before the episodic memory test (retrieval-stereotype condition), and a control group where stereotypes were not explicitly activated. Because we initially were interested in retrieval effects, we first alternated assignment to the control and the retrieval-stereotype groups, and then tested participants in the encoding-stereotype group. All participants were tested within a 6-month period, using the same recruitment and screening procedures, and all received the same monetary compensation for their participation.

A series of ANOVAs comparing all three groups revealed no group difference in age (controls = 74.1, encoding-stereotype = 76.2, retrieval-stereotype = 75.1), general cognitive functioning on the MMSE (means = 28.3, 28.0, and 27.7), or depressive symptoms on the GDS (means = 4.71, 4.06, 3.85), all $p$'s < .30. There was a significant group difference in years of education, $F(2, 122) = 5.12$, MSE = 10.956, $p = .007$, $\eta_p^2 = .077$, as the encoding-stereotype group had fewer years of education than the retrieval-stereotype group (16.0 vs 18.3), $t\ (81) = 2.95$, SEM = .786, $p = .004$, Cohen's $d = .66$, although neither group reliably differed from the control group (17.2). Education data were not recorded for one encoding-stereotype participant. Because of this difference, education was included as a covariate in the group comparisons reported in the Results section.

### Tasks and Procedures

#### Stereotype manipulation

We used the same stereotype manipulation as in Wong and Gallo (2016). Older adults in the stereotype conditions read a passage containing scientific findings of age-related decline in memory, word-finding, and multitasking, whereas older adult controls read an age-neutral passage on language. These passages were presented as a reading comprehension task. In addition, we reiterated key ideas from the relevant passage throughout the testing phase, so that the stereotype information would remain active (see Procedure section).

#### Episodic memory task

During the encoding phase of the source recollection task, participants were visually presented with two different lists of words that depicted common objects (e.g., doll, tie). To minimize the unintentional activation of stereotype threat, none of the participants were told prior to the encoding phase that their memory would be tested. Instead, we had participants make semantic judgments. For the first list, they were presented with 40 words and judged whether the item was pleasant (yes/no, self-paced). For the second list, participants were presented with 40 words and judged whether the item is typically made in a factory (yes/no, self-paced). Half of the words in this second list had also been presented in the first list, whereas half had not, so that the relationship between the two lists was not mutually exclusive. This procedure was designed to make the subsequent recollection test more challenging (i.e., to minimize a "recall-to-reject" strategy, cf. Gallo, Cotel, Moore, & Schacter, 2007).

During the test phase of this task, participants had to indicate whether each of the test items had previously appeared in the factory list (yes/no, self-paced), regardless of whether the item was presented in the pleasantness list, followed by a confidence rating (0 = guess, 1 = low, 2 = medium, 3 = high). Importantly, both factory and pleasantness items should have been familiar at test, thereby encouraging participants to recollect the earlier judgment to make their decision, rather than relying on a more vague sense of familiarity. There were 80 test items presented in a randomized order: 20 items that had been presented in both the pleasantness and factory lists, 20 items that had only been presented in the factory list, 20 items that had only been presented in the pleasantness phase, and 20 items that had not been studied before. Test items were randomly presented, and for counter-balancing, each item was rotated through each item-type condition across participants. After the test, participants indicated whether they expected a memory test and if so, they reported any strategies that they used during encoding.

#### General knowledge task

We used a trivia task based on the Nelson and Narens (1980) norms to assess general knowledge or semantic memory. There were 60 trials, drawn from three levels of difficulty reported in Dodson, Bawa, & Krueger (2007): 20 easy (.81 correct), 20 medium (.44), and 20 difficult (.07). All trivia questions had one-word answers about history, geography, popular culture, sports, or definitions. Participants were asked to write down their answer and were encouraged to guess if they were not sure. The trivia score was the total number of correctly recalled answers.

#### Working memory task

We used a backward digit span task with a concurrent imagery task (e.g., Lefebvre, Marchand, Eskes, & Connolly, 2005; see Conway et al., 2005 for a review) to measure working memory. Participants heard the names of three objects and formed mental images to keep in mind. Next, they heard a string of numbers at the rate of 1 s per number and had to orally repeat the numbers back in the reverse order. The string size started at two and increased by one until the participant missed two strings in a row. The backward digit span score was the total number of strings recalled correctly. There were two trials per string size, starting from set size two and progressing to eight (i.e., 14 strings total). After completing the backward digit span, they recalled the names of the three mental imagery objects, and then reported any strategies they used.

#### Metacognitive assessment

We developed a metacognitive insight questionnaire (MIQ; Supplementary Appendix A) to assess global metacognition. Participants rated their general abilities in three domains relative to when they were 20 years old: attention/cognition (working memory), short-term memory (episodic memory), and language (semantic memory). The 7-point scale ranged from much worse (−3) to same (0) to much better (+3). Participants completed the MIQ prior to the memory tasks (pretask) and again after the memory tests (post-task). Local or task-based metacognition was assessed using performance estimates immediately before and after each test. For these judgments, participants predicted how well they would do on that particular test (pretest) or how well they actually did (post-test) relative to a 20-year-old self, using the same scale as the MIQ. Before making pretest judgments, participants received the test instructions and completed one practice trial.

### Procedure

The order of all tasks is depicted in Figure 1. In the encoding-stereotype group, participants first read the stereotype passage, then completed the episodic memory task encoding phase, the pretask MIQ, the episodic memory test phase, the general knowledge task, the working memory task, and the post-task MIQ. All tasks were presented in this order, with the exception that the order of the general knowledge and working memory tasks was counterbalanced across participants. Identical procedures were used for the retrieval-stereotype group and the control group, except the stereotype/neutral control passage was read immediately after the episodic memory encoding phase. To keep the explicit stereotype active, we read excerpts from the stereotype (or control) paragraph that were relevant to each test domain (episodic memory, general knowledge,

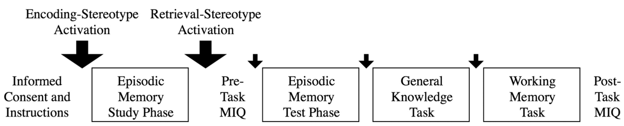


**Figure 1.** Experimental task order. Stereotypes were activated before or after encoding on the episodic memory task (large arrows), and also were reiterated just prior to each cognitive test (small arrows). The order of the working memory and general knowledge tasks was counterbalanced across participants in each stereotype group. MIQ = Metacognitive insight questionnaire.

and working memory) just prior to taking each test. This information was presented as part of the task instructions (i.e., we told participants that each task was designed to test the ideas that were presented in the original scientific passage).

Once participants completed the post-task MIQ, they received a multiple-choice test to assess comprehension of the stereotype/neutral passage. Then, they described their lifestyle when they were 20 years old and whether they experienced anxiety during the study. Finally, all of the psychometrics and demographics forms (including date of birth) were administered at the end of the experiment. No other tasks or measures were administered in this study.

## Results

All results were considered significant at the conventional level ($p < .05$), and all group comparisons using analysis of covariance (ANCOVA) included education as a covariate.

### Episodic Memory Task

#### Encoding phase

There were no reliable differences between the three stereotype conditions on the average number of "yes" responses to the pleasantness or factory judgments, nor were there any reliable effects of stereotype condition on associated response latencies (all $p$'s > .10). On the post-test questionnaire, approximately 25% of the participants indicated they might have expected a memory test during the encoding phase ($n$ = 11 in the control group, $n$ = 12 in the encoding-stereotype group, and $n$ = 6 in the retrieval-stereotype group), but comparison of these 29 participants to the other 97 revealed no differences on either of the two memory accuracy scores described below (both $t$'s < 1). All participants were included in the analyses reported below.

#### Retrieval phase

Hit and false alarm rates are presented in Table 1. Across conditions, participants were more likely to correctly endorse items studied in both lists compared to items studied only in the factory list (critical targets), these hit rates were greater than false alarms to items studied in the

**Table 1.** Mean Proportion of Source Recollection Test Items Endorsed With a "Yes" Response as a Function of Stereotype Condition

| | Stereotype at encoding | Stereotype at retrieval | Control |
|---|---|---|---|
| Both list hits | 0.78 (0.04) | 0.76 (0.03) | 0.72 (0.03) |
| Factory list hits | 0.64 (0.03) | 0.63 (0.03) | 0.61 (0.03) |
| Pleasant list FAs | 0.56 (0.04) | 0.47 (0.05) | 0.48 (0.04) |
| Nonstudied FAs | 0.10 (0.02) | 0.14 (0.03) | 0.19 (0.04) |

*Note*: Parentheses report standard error of each mean. FA = false alarms.

pleasantness list (studied lures), which in turn were greater than false alarms to nonstudied lures (all *p*'s < .001). Thus, participants were quite good at differentiating between critical targets and nonstudied lures, and they also were able to use recollection to successfully differentiate between critical targets and studied lures, each of which should have been familiar. However, by design, this source discrimination was challenging and participants made many false alarms to lures that had been studied in the pleasantness list, relative to nonstudied lures.

To facilitate comparisons, we calculated two accuracy measures to control for possible shifts in overall response criteria between the groups (Figure 2). The *source discrimination score* (factory target hits minus pleasantness lure false alarms) measured participants' ability to use recollection to discriminate between the critical targets and the studied lures, as each of these item types had been studied once and hence should have been familiar. The *source confusion score* (pleasantness lure false alarms minus nonstudied lure false alarms) measured the effect of studying an item in the nontarget list on memory errors, using nonstudied lures as a baseline. This measure assesses susceptibility to source confusions during recollection attempts.

An ANCOVA comparing source discrimination scores across the three groups was not significant ($F < 1$), but there was a significant group effect on source confusions, $F(2, 121) = 3.53$, MSE = .066, $p = .03$, $\eta_p^2 = .055$. The encoding-stereotype group had higher source confusion scores than controls, $F(1, 80) = 6.69$, MSE = .065, $p = .01$, $\eta_p^2 = .077$, whereas the retrieval-stereotype group did not differ from controls ($F < 1$). Source confusions also were numerically greater in the encoding-stereotype group compared to the retrieval-stereotype group, although this difference failed to reach significance, $F(1, 80) = 2.62$, MSE = .072, $p = .11$, $\eta_p^2 = .032$ (but see next section). There were no reliable effects of stereotype condition on response latencies to

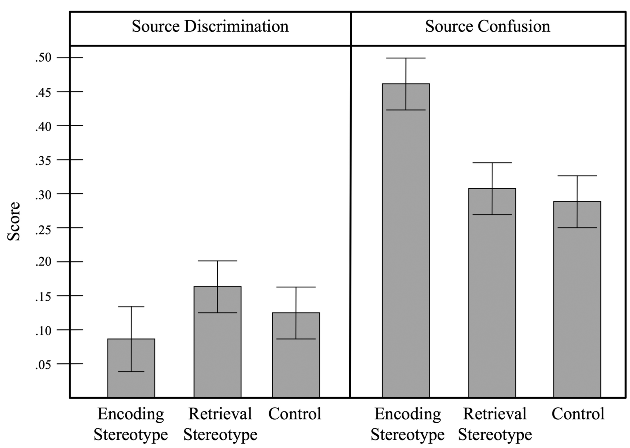


**Figure 2.** Source discrimination scores (criterial target hits—studied lure FAs) and source confusion scores (studied lure FAs—nonstudied lure FAs) on the episodic memory test. Bars report standard error of each mean. FA = false alarms.

correct responses (all *p*'s > .09). Overall, these data revealed that explicitly activating stereotype information at encoding elevated recollection confusions.

### High confidence responses

Across all groups, participants were very confident in their correct responses, with mean confidence = 2.45 out of 3 for hits (0.55 *SD*), and 2.43 out of 3 for correct rejections of studied lures (0.57 *SD*), with no reliable differences between the groups in overall confidence. This tendency for high-confidence responding suggests that participants had relied primarily on recollection at test, although familiarity may still have contributed to performance. To further isolate recollection-based responding, we re-analyzed the recollection test data considering only those responses that were made with medium or high confidence, under the assumption that these responses would be most likely to reflect a strong sense of recollection (cf. Gallo & Roediger, 2002). Data from 2 encoding-stereotype participants was excluded from this analysis because they did not have high-confidence responses. There was no effect of group on source discrimination scores, $F < 1$, but there was a significant effect on source confusion scores, $F(2, 119) = 3.87$, MSE = .082, $p = .02$, $\eta_p^2 = .061$. Source confusion scores were greater in the encoding-stereotype group (.49) compared to the control group (.35), $F(1, 78) = 5.48$, MSE = .084, $p = .02$, $\eta_p^2 = .066$, and also compared to the retrieval-stereotype group (.32), $F(1, 78) = 5.44$, MSE = .081, $p = .02$, $\eta_p^2 = .065$. There was no difference between the control group and the retrieval-stereotype group ($F < 1$). These data suggest that activating the stereotype at encoding elevated source confusions by increasing false recollection.

## Metacognitive Reports

We measured global metacognition using the MIQ, collecting self-report ratings of one's general cognitive abilities relative to their younger self (Table 2). We also measured task-specific or local metacognition (Table 3), but there were no reliable effects of the stereotype manipulation on these judgments and so they are not discussed further. For the MIQ data, there were no significant differences between the two stereotype groups on average ratings for any of the domains for either of the MIQ administrations (all *p*'s > .05), and for simplicity, we report MIQ analyses comparing the older adults who received stereotype activation at retrieval to controls. This focus offers the purest analysis of stereotype effects on metacognitive assessment, as the MIQ immediately followed the stereotype manipulation only in the retrieval condition.

A 3 (MIQ domain) × 2 (stereotype condition) ANCOVA on the pretask MIQ ratings revealed a main effect of stereotype, $F(1, 81) = 4.69$, MSE = 2.26, $p = .03$, $\eta_p^2 = .055$, and a significant interaction between stereotype and domain, $F(2, 162) = 3.96$, MSE = .309, $p < .02$, $\eta_p^2 = .047$. Older adults in the stereotype condition rated themselves lower than

**Table 2.** Mean Pretask and Post-task Global Metacognitive Ratings on the MIQ Relative to a 20-Year-old Self (from −3 to +3, with 0 = same) in Three Cognitive Domains as a Function of Stereotype Condition

| | | Stereotype at encoding | Stereotype at retrieval | Control |
|---|---|---|---|---|
| Pretask | | | | |
| | Episodic memory | -0.66 (0.10) | -0.82 (0.13) | -0.70 (0.15) |
| | Working memory | -0.57 (0.16) | -0.82 (0.16) | -0.22 (0.17) |
| | General knowledge | -0.30 (0.10) | -0.05 (0.13) | +0.40 (0.16) |
| Post-task | | | | |
| | Episodic memory | -0.89 (0.13) | -0.83 (0.14) | -0.72 (0.13) |
| | Working memory | -0.92 (0.13) | -0.74 (0.15) | -0.55 (0.15) |
| | General knowledge | -0.57 (0.16) | -0.27 (0.13) | +0.02 (0.17) |

*Note*: Parentheses report standard error of each mean. The encoding group received stereotype activation, then the episodic memory study phase, then the pretask MIQ, whereas the retrieval group received the episodic memory study phase, then stereotype activation, then the pretask MIQ. Both groups took the post-task MIQ at the end of the secondary tasks (see Figure 1). MIQ = Metacognitive insight questionnaire.

**Table 3.** Mean Pretask and Post-task Local Metacognitive Ratings Relative to a 20-Year-old Self (from −3 to +3, with 0 = same) for the Three Cognitive Tasks as a Function of Stereotype Condition

| | | Stereotype at encoding | Stereotype at retrieval | Control |
|---|---|---|---|---|
| Pretask | | | | |
| | Episodic memory | -0.31 (0.24) | -0.50 (0.22) | -0.38 (0.24) |
| | Working memory | -1.10 (0.20) | -1.31 (0.20) | -1.26 (0.19) |
| | General knowledge | -0.52 (0.22) | -0.83 (0.23) | -0.12 (0.23) |
| Post-task | | | | |
| | Episodic memory | -0.48 (0.20) | -0.98 (0.20) | -0.83 (0.19) |
| | Working memory | -0.95 (0.20) | -1.05 (0.23) | -1.07 (0.21) |
| | General knowledge | -0.90 (0.25) | -0.74 (0.24) | -0.21 (0.27) |

*Note*: Parentheses report standard error of each mean.

controls on working memory, $F(1, 81) = 7.03$, MSE = 1.13, $p = .01$, $\eta_p^2 = .08$, and on semantic memory, $F(1, 81) = 5.14$, MSE = .906, $p = .03$, $\eta_p^2 = .06$. Activating stereotypes about aging-related decline were not expected to reduce beliefs about one's semantic memory abilities. However, unlike the other cognitive domains on the MIQ, the items for semantic memory primarily emphasized language abilities. Thus, the group difference observed on these items might be due to a stereotype lift effect in controls, because the control passage emphasized a high degree of language proficiency in all people. The stereotype effect on episodic memory was not reliable ($F < 1$), even though three of the four relevant comparisons from Table 2 show a numerical effect in the predicted direction.

These results indicate that the stereotype tended to make people rate their own personal aging decline as more severe in several cognitive domains, but there were two caveats to this conclusion. First, a mediation analysis using the PROCESS bootstrapping procedure (Hayes, 2013), found no evidence that these stereotype effects on the first administration of the MIQ mediated stereotype effects on source confusion scores. Second, analysis of the second administration of the MIQ (i.e., after the cognitive tasks had been completed) yielded no effect of stereotype or interaction (both $p$'s > .21), potentially because stereotype effects on global metacognition declined after several cognitive tasks were administered.

## Secondary Tasks: Working Memory and General Knowledge

Data from the secondary tasks are reported in Table 4. Overall there was little evidence that stereotypes impacted these tasks, although as discussed, it is possible that the stereotype effect may have faded by the time these tasks were administered. On the working memory task, participants were good at recalling the mental images ($M = 2.60$ out of 3) and there was no group difference on this aspect of the task ($F < 1$). With respect to the number of digit strings correctly recalled in backward order, an ANCOVA revealed no significant effect of group, $F(2, 121) = 2.03$, MSE = 5.87, $p = .14$, $\eta_p^2 = .032$. On the General Knowledge task, across all conditions, performance reliably tracked item difficulty (mean proportion of correct trivia questions = .72 for easy questions, .54 for medium difficulty, and .25 for high difficulty). A 3 (stereotype condition) × 3 (item difficulty) ANCOVA yielded an effect of item difficulty, $F(2, 242) = 10.44$, MSE = .013, $p < .001$, $\eta_p^2 = .079$, but no effect of stereotype and no interaction (both $F$'s < 1).

**Table 4.** Mean Performance on Secondary Cognitive Tasks as a Function of Stereotype Condition

| | | Stereotype at encoding | Stereotype at retrieval | Control |
|---|---|---|---|---|
| Working memory | | | | |
| | Images | 2.51 (0.12) | 2.71 (0.09) | 2.60 (0.12) |
| | Digits | 6.26 (0.36) | 7.74 (0.40) | 7.43 (0.39) |
| General knowledge | | | | |
| | Easy | 0.67 (0.04) | 0.73 (0.03) | 0.76 (0.03) |
| | Medium | 0.50 (0.05) | 0.56 (0.04) | 0.57 (0.04) |
| | Hard | 0.20 (0.03) | 0.26 (0.03) | 0.27 (0.03) |

*Note*: The stereotype manipulation always occurred prior to these secondary tasks. Parentheses report standard error of each mean.

### Post-task Questionnaire

Overall participants did quite well on the test assessing comprehension of the stereotype/neutral passage, with slightly greater comprehension of the neutral passage (mean = 4.4 out of 5) than the stereotype passage (mean = 4.0 in the encoding group, and 4.2 in the retrieval group), and no difference between the two stereotype groups, $t$ (81) =1.06, SEM = .23, $p$ = .29. With respect to the anxiety question, anxiety was reported by 59% of the participants in the control group, 33% in the encoding-stereotype group, and 43% in the retrieval-stereotype group. Thus, the explicit activation of stereotypes did not increase the number of participants reporting anxiety.

## Discussion

Explicitly activating negative aging stereotypes prior to encoding increased source recollection confusions in older adults, but the same stereotype manipulation had no effect when activated just prior to retrieval. These results are broadly consistent with prior literature showing that the explicit activation of stereotype threat prior to encoding can impair episodic memory performance in older adults (e.g., Barber & Mather, 2013b; Hess et al., 2003; Hess et al., 2009; Hess et al., 2009; Hess & Hinson, 2006; Kang & Chasteen, 2009; Mazerolle et al., 2012), and they more specifically show that stereotype activation can increase false recollection, operationalized as memory errors made with medium to high confidence on a source recollection task. By contrast, stereotype activation at retrieval did not impact memory as strongly as stereotype activation at encoding, and this may partially explain why previous studies of stereotype threat effects at retrieval have produced mixed results (Smith et al., in press; Thomas & Dubois, 2011; Wong & Gallo, 2016).

One interpretation of our recollection task results is that stereotype threat primarily impacts encoding processes. However, it is unlikely that stereotype threat affected explicit memorization strategies, because our participants were not told that their memory would be tested during encoding, and the speed or quality of the semantic judgments at encoding did not differ across our stereotype conditions. Instead, activating the stereotype at encoding might have diminished incidental thoughts about the studied words and semantic judgments during encoding, focusing participants on more stereotype-relevant thoughts. This distraction from the task might have reduced the formation of associations between studied words and semantic judgments during encoding, associations that would otherwise benefit retrieval monitoring processes during recollection attempts. Activating stereotypes prior to encoding also might have caused participants to lose confidence in the quality of their encoding. These hypothetical mechanisms of stereotype threat ultimately could have impaired retrieval monitoring in the stereotype-encoding group, thereby increasing false recollection.

To our knowledge, no prior stereotype threat study in older adults has used a source recollection test, and there were two other findings from our task that are important to highlight. First, comparing the encoding-stereotype and control groups, we found that source confusion scores were more sensitive to stereotype effects than source discrimination scores. This difference may have been driven by the relatively high source confusions in our task, which yielded high source confusion scores and low source discrimination scores even in the control group, potentially making the former measure more sensitive to experimental manipulations such as stereotype activation. Second, the increase in source confusions scores that we observed was driven by an increase in false alarms to studied lures as well as a concurrent decrease in false alarms to nonstudied lures. The latter effect is analogous to the stereotype-induced decrease in false alarms to nonstudied lures observed by Barber and Mather (2013b). Barber and Mather (2013b) argued that explicitly activating stereotype threat at encoding made participants respond more conservatively at test, and our results are not inconsistent with this idea. Stereotype activation may have increased the subjective experience of false recollection and recollection confusions, due to the mechanisms described above, while at the same time making participants respond more conservatively.

Our findings differ from those of Krendl et al. (2015), who found that subliminal stereotype activation at retrieval increased false recognition errors in older adults, with less robust effects of stereotype activation at encoding. While speculative, it may be the case that the explicit stereotype activation procedure used here is more likely to preoccupy participants with aging-related worries or ruminations that can impact memory encoding processes, whereas the subliminal procedure used in Krendl et al. primarily introduces a response bias during memory testing. Future work will need to directly test this idea, ideally by comparing subliminal and explicit stereotype activation procedures using the same cognitive tasks and sampling from the same older adult population. A reviewer wondered whether our administration of the metacognitive assessment just prior to the memory test diffused the stereotype manipulation in the retrieval condition. While we cannot rule out this possibility, both Bouazzaoui et al. (2016) and Wong and Gallo (2016) also administered a metacognitive assessment just prior to their stereotype manipulation (the former at encoding, the later at retrieval), and a significant stereotype effect was found in each of those studies. Thus, stereotype effects can be obtained following metacognitive assessments.

In addition to finding a stereotype effect on episodic memory, we also found that stereotype activation caused individuals to rate themselves as having greater aging-related decline in several cognitive domains on the MIQ. This effect conceptually replicates that observed by Bouazzaoui et al. (2016), suggesting that stereotype threat diminishes one's sense of personal memory ability. However, unlike Bouazzaoui et al. (2016), these stereotype effects on metacognitive ratings did not mediate the effects of stereotype activation on episodic memory. We also did not find stereotype effects on task-based metacognitive ratings. Thus, we found no evidence that stereotype effects on metacognitive beliefs about one's own abilities could, in turn, impair task performance. One possible explanation for this difference is that our MIQ assessed cognitive decline with aging, whereas Bouazzaoui et al. (2016) assessed memory complaints as well as memory self-efficacy. Memory complaints and beliefs about self-efficacy may have a more direct impact on task motivations and strategies than the metacognitive measure that we used.

We found that stereotype activation impacted source recollection errors and metacognitive ratings, but stereotype activation had minimal impact on our secondary tasks of general knowledge and working memory. To our knowledge, no prior study has investigated stereotype threat effects on general knowledge tasks in older adults, but because this kind of crystallized intelligence is not usually affected by aging or portrayed in negative aging stereotypes, we did not expect stereotypes to impact this task. With respect to working memory, two previous studies have found significant effects of stereotype activation on working memory tasks in older adults (Barber & Mather, 2013a; Mazerolle et al., 2012), whereas at least two other studies have not (Hess et al., 2009; Popham & Hess, 2015a). Our study also did not find an effect on working memory, although it should be reiterated that stereotype threat might have diminished by the time these secondary tasks were administered.

In conclusion, the primary new finding from this study was that explicitly activating aging stereotypes at encoding, but not retrieval, significantly increased false recollection in older adults. These results are broadly consistent with previous studies showing that explicitly activating stereotypes prior to encoding can impair memory performance in older adults (e.g., Hess et al., 2003; Hess et al., 2009), and they extend this finding to high-confidence errors on a source recollection task. This finding, along with the lack of an effect of stereotype activation at retrieval, is consistent with the hypothesis that explicitly activating stereotype threat can reduce the effectiveness of episodic memory encoding, thereby making it more difficult to accurately monitor memory errors at retrieval.

## Supplementary Material

Supplementary data is available at *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences* online.

## Funding

This work was supported by an American Psychological Association Dissertation Research Award to J. T. Wong, and by the National Institute of Aging of the National Institutes of Health (R21AG049931). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Conflict of Interest

None reported.

## References

Barber, S. J. (2017). An examination of age-based stereotype threat about cognitive decline: Implications for stereotype threat research and theory development. *Perspectives in Psychological Science*, **12**, 62–90.

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

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

Barber, S. J., & Mather, M. (2013b). Stereotype threat can reduce older adults' memory errors. *Quarterly Journal of Experimental Psychology (2006)*, **66**, 1888–1895. doi:10.1080/17470218.2013.840656

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 (Hove, England)*, **24**, 659–668. doi:10.1080/09658211.2015.1040802

Brink, T. L., Yesavage, J. A., Lum, O., Heersema, P., Adey, M. B., & Rose, T. L. (1982). Screening tests for geriatric depression. *Clinical Gerontologist*, **1**, 37–44.

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

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, **12**, 769–786.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, **104**, 263–294.

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

Dodson, C. S., Bawa, S., & Krueger, L. E. (2007). Aging, metamemory, and high-confidence errors: A misrecollection account. *Psychology and Aging*, **22**, 122–133. doi:10.1037/0882-7974.22.1.122

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, **12**, 189–198.

Gallo, D. A., & Roediger, H. L. III. (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory and Language*, **47**, 469–497.

Gallo, D. A., Cotel, S. C., Moore, C. D., & Schacter, D. L. (2007). Aging can spare recollection-based retrieval monitoring: The importance of event distinctiveness. *Psychology and Aging*, **22**, 209–213. doi:10.1037/0882-7974.22.1.209

Grady, C. (2012). The cognitive neuroscience of ageing. *Nature Reviews. Neuroscience*, **13**, 491–505. doi:10.1038/nrn3256

Hayes, A. F. (2013). *Introduction to mediation, moderation and conditional process analysis*. New York, NY: Guilford Press.

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

Hess, T. M., Emery, L., & Queen, T. L. (2009a). Task demands moderate stereotype threat effects on memory performance. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, **64B**, 482–486. doi:10.1093/geronb/gbp044

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

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

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, **88**, 67–85.

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

Krendl, A. C., Ambady, N., & Kensinger, E. A. (2015). The dissociable effects of stereotype threat on older adults' memory encoding and retrieval. *Journal of Applied Research in Memory and Cognition*, **4**, 103–109. doi:10.1016/j.jarmac.2015.02.001

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

Lefebvre, C. D., Marchand, Y., Eskes, G. A., & Connolly, J. F. (2005). Assessment of working memory abilities using an event-related brain potential (ERP)-compatible digit span backward task. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, **116**, 1665–1680. doi:10.1016/j.clinph.2005.03.015

Levy, B. R., Zonderman, A. B., Slade, M. D., & Ferrucci, L. (2012). Memory shaped by age stereotypes over time. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, **67**, 432–436. doi:10.1093/geronb/gbr120

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

Nelson, T.O., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior*, **19**, 338–368.

Popham, L. E., & Hess, T. M. (2015a). Age differences in the underlying mechanisms of stereotype threat effects. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, **70**, 225–234. doi:10.1093/geronb/gbt093

Popham, L. E., & Hess, T. M. (2015b). Stereotype threat. In S. Krauss (Ed.), *The encyclopedia of adulthood and aging* (pp. 1–5). West Sussex: John Wiley & Sons.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 803–814.

Smith, A. M., Gallo D.A., Barber, S. J., Maddox K., & Thomas, A. K. (in press). Stereotypes, warnings, and demographic characteristics influence older adults' susceptibility to associative false memory errors. *The Gerontologist*.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, **69**, 797–811.

Thomas, A. K., & Dubois, S. J. (2011). Reducing the burden of stereotype threat eliminates age differences in memory distortion. *Psychological Science*, **22**, 1515–1517. doi:10.1177/0956797611425932

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