Education and Retirement Status Moderate Older Adults' False Memory Susceptibility Under Stereotype Threat

A thesis submitted by

Amy M. Smith

in partial fulfillment of the requirements for the degree of

Master of Science

in

Psychology

Tufts University

February 2016

Adviser:

Dr. Ayanna K. Thomas

Abstract

One study examining the effects of stereotype threat on older adults' false memory susceptibility reported threat-related *increases* in false memories, whereas another study reported *decreases*. Both studies used the Deese-Roediger-McDermott (DRM) paradigm but differed in whether or not they provided participants with a warning about DRM errors. We examined whether this warning is the key factor in determining whether threat increases or decreases false memory susceptibility. We further evaluated the influence of individual differences in moderating stereotype threat susceptibility. In comparison to those who were not threatened or warned, older adults under stereotype threat demonstrated increased susceptibility to false memories. However, this effect only emerged for individuals who were retired and/or highly educated. This suggests that the effects of threat on false memory susceptibility in the DRM paradigm may be contingent on one's identification with the threatened group, rather than the presence or absence of a warning.

Table of Contents

List of Tables

List of Figures

Introduction

Slow thinking, senility, and forgetfulness are just a few of the negative stereotypes about the elderly that are common in Western culture. When older adults (i.e., 55 and older) internalize such negative beliefs they have been shown to experience *stereotype threat.* This phenomenon is characterized by older adults demonstrating sub-standard performance on a wide range of cognitive tasks after being confronted with stereotypes about aging. For example, older adults who read passages emphasizing the negative effects of aging on cognition subsequently recalled fewer words than those who read positive aging passages (Hess, Auman, Colcombe, & Rahhal, 2003). Similarly, older adults under stereotype threat have shown sub-standard working memory performance (Mazerolle, Régner, Morisset, Rigalleau, & Huguet, 2012), word recall (Chasteen, Bhattacharyya, Horhota, Tam, & Hasher, 2005; Hess & Hinson, 2006; Hess, Hinson, & Hodges, 2009), photo recall (Stein, Blanchard-Fields, & Hertzog, 2002), recall of prose passages (Kang & Chasteen, 2009), and word recognition (Hess, Emery, & Queen, 2009). Stereotyped older adults may experience anxiety about the possibility of behaving in a stereotype-consistent manner, thus resulting in the observed performance decrements (Schmader, Johns, & Forbes, 2008).

More recently, Thomas and Dubois (2011) found that older adults under stereotype threat demonstrated greater susceptibility to false memories in the DRM paradigm as compared to those under reduced threat. Thomas and Dubois (2011) hypothesized that this false memory increase occurred because threat placed a burden on older adults' cognitive resources, leaving fewer resources available when attempting to remember studied material at test. Stereotype threat has been shown to reduce older adults' working memory capacity, providing support for the resource depletion mechanism (Mazerolle et al., 2012).

Whereas Thomas and Dubois (2011) found increased false recognition for older adults under threat, just the opposite was demonstrated by Wong and Gallo (2015). Wong and Gallo (2015) suggested that the reduction in false memories under threat may have resulted from the combination of threat with a warning about the deceptive nature of the DRM paradigm, which they provided to their participants. Thus, in the context of the same methodology, research has yielded conflicting results and consequentially, provided support for different underlying mechanisms. The goal of the present experiment was to address this discrepancy.

Stereotype threat is a social evaluation phenomenon that should be most influential for individuals who have internalized the negative stereotype. The threat manipulation used in the two previous studies was designed to target individuals who identify as "older adults" and who value their cognitive abilities. However, neither of these studies considered the influence of identity-related measures. Such measures may have helped inform researchers as to which older adults identified with the stigmatized group, and, by extension, which older adults experienced the negative effects of threat. In the present study, we examined the individual and combined effects of stereotype threat and the DRM warning on false recognition, while also considering the moderating influence of measures related to identification with the targeted group.

Stereotype Threat and Executive Resource Depletion

Thomas and Dubois (2011) argued that stereotype threat exacerbated age-related false memory susceptibility because threat resulted in the depletion of cognitive resources. This explanation was derived from the *executive resource depletion* account of stereotype threat, which is a widely adopted explanation for why stereotype threat negatively affects cognitive functioning (e.g., Beilock, Rydell, & McConnell, 2007; Régner et al., 2010; Schmader & Johns, 2003). This model is based on the premise that, at any given moment, we have limited mental (i.e., *executive*) resources available for performing cognitive tasks. Stereotype threat consumes some of these resources via active self-monitoring of one's performance during the stereotyped task, self-regulatory efforts aimed at suppressing negative thoughts associated with threat, and/or a physiological response to threat that impairs prefrontal processing (Schmader, Johns, & Forbes, 2008; for a review see Schmader, 2010). These disruptions leave an individual with fewer resources available to perform the given task. As a result, cognitive performance suffers for those under threat.

In a typical DRM experiment, participants are presented with wordlists, each containing the 15 strongest semantic associates (e.g., *bed, rest, tired*) of a non-presented lure word (e.g., *sleep*). The common finding is that participants often falsely remember the non-presented lure word (Roediger & McDermott, 1995, Stadler, Roediger, & McDermott, 1999). Research suggests that these false memories can be avoided when participants engage in effective source monitoring. That is, according to the *activation-monitoring framework*, false memories in the DRM paradigm arise because the non-presented lure word is *internally* activated during list learning, and participants inaccurately attribute this memory to an *external* source (i.e., the studied lists) (McDermott & Watson, 2001; but see also Roediger, Balota, & Watson, 2001). At test, if participants carefully monitor the source of their memories for words, they will be more likely to remember which specific items were learned and which were merely products of their own mental processes. False memories may also be avoided when participants remember item-specific information that can be used to differentiate between previously encoded and internally generated stimuli (*fuzzy trace theory*; Brainerd & Reyna, 1990, 2002; Reyna & Brainerd, 1995). These theoretical frameworks suggest that reducing or eliminating false memories in the DRM paradigm demands conscious effort and requires cognitive resources.

Consistent with this idea, when young adult participants are warned about the deceptive construction of DRM lists, they sometimes show reductions in false memories (e.g., Gallo, Roediger, & McDermott, 2001; McCabe & Smith, 2002). Warnings may encourage conscious source monitoring (Thomas, Bulevich, & Chan, 2010; Thomas & Bulevich, 2006) and/or item-specific recollection (Gallo et al., 2001). In comparison to young adults, older adults demonstrate source-monitoring deficits (e.g., Dehon & Brédart, 2004, McIntyre & Craik, 1987) and tend to over-rely on gist, as opposed to item-specific, processing (Thomas & Sommers, 2005; Tun, Wingfield, Blanchard, & Rosen, 1998). Older adults may therefore be more susceptible to the DRM false memory effect because of source-monitoring deficits and/or because of reduced availability/accessibility to item-specific information. This may explain why warnings have been unsuccessful at reducing the number of false memory errors that older adults make (McCabe & Smith, 2002).

The executive resource depletion account coheres nicely with both the activation-monitoring and fuzzy trace theories of DRM false memory susceptibility. Both source monitoring and item-specific recollection are mentally effortful processes that require executive resources. When stereotype threat depletes these limited resources, efforts aimed at careful source monitoring and item-specific remembering may be blunted during a DRM recognition or recall test. As a result, threatened individuals may experience increased false memory susceptibility. As previously mentioned, older adults over-rely on gist-based processing and are worse at source monitoring than young adults. Thus, stereotype threat may compound the issue by further depleting their cognitive resources, which may then increase their reliance on gist response strategies and/or worsen their source monitoring capabilities. Thomas and Dubois

(2011) provided evidence for this supposition since threatened older adults falsely remembered related lure words more often than those who were not under threat.

The Regulatory Fit Hypothesis

Alternatively to the executive resource depletion account, stereotype threat researchers have provided support for the *regulatory fit hypothesis* (Barber & Mather, 2013; Chalabaev, Major, Sarrazin, & Cury, 2012; Grimm, Markman, Maddox, & Baldwin, 2009; Maddox, Baldwin, & Markman, 2006; Wong & Gallo, 2015). This hypothesis maintains that people perform better on cognitive tasks when their internal regulatory state (i.e., promotion or prevention focused) matches the task's reward structure (i.e., gains-based or losses-based). People with a promotion focus are concerned with the presence or absence of gains, so a regulatory fit is achieved when they encounter a task in which they receive rewards (e.g., gaining money for correct answers). People with a prevention focus are concerned with the presence or absence of losses, so these individuals achieve a regulatory fit when they must avoid punishment on a given task (e.g., losing money for wrong answers). In the context of their study, Wong and Gallo (2015) hypothesized that stereotype threat induced a prevention focus because older adults wanted to avoid making mistakes on the recognition test that would confirm the negative stereotype (see also Barber & Mather, 2013; Seibt & Förster, 2004). Furthermore, warning participants about the deceptive nature of the DRM paradigm may have turned the recognition test into a losses-based task by encouraging individuals to *avoid* making false memory errors. Therefore, they reasoned that a regulatory fit was achieved when stereotype threat activated a prevention focus and the DRM warning provided a losses-based task at test. This resulted in the observed reduction in false memory susceptibility for older adults who were threatened and warned. These findings are consistent with a previous study in which a regulatory fit was demonstrated when negative stereotyping led to better math performance on a test that emphasized losing points for incorrect responses in comparison to a test that emphasized gaining points for correct responses (Grimm et al., 2009).

Whether executive resource depletion or regulatory fit more accurately explains how stereotype threat affects older adults' cognition is a debate that goes beyond the findings in the context of the DRM paradigm. Older adults have demonstrated reduced working memory capacity when under stereotype threat, providing support for executive resource depletion (Mazerolle et al., 2012). Alternatively, older adults under threat have demonstrated better recall performance when told that they would lose money for each word forgotten than when they were told that they would receive monetary gains for correct recall (Barber & Mather, 2013), providing support for the regulatory fit hypothesis. Furthermore, this experiment directly contradicted the resource depletion account as threatened older adults demonstrated no difference in their pre- and post-threat working memory performance. Hess et al. (2009) also found that stereotype threat did not affect older adults' performance on a computation span working memory task. Threatened older adults in this study also demonstrated conservative responding on a test of free recall, which could have been indicative of their adoption of a prevention focus.

Clearly this theoretical dichotomy warrants further exploration. The DRM paradigm is a useful tool for this task, given that it allows us to examine the effects of threat on both veridical and false memory. By doing so, we will be able to compare our findings to previous research that measured accurate memory performance (e.g., Hess et al., 2009), as well as the growing body of literature examining false memory (Thomas & Dubois, 2011; Wong & Gallo, 2015).

Identity and Stereotype Threat

The extent to which an individual identifies with a stigmatized group plays a role in determining whether cognitive performance is affected by stereotype threat. In fact, merely requiring Black students to record their race before taking a verbal test lead to poorer performance than that of Black students who did not report their race (Steele & Aronson, 1995). Since race is a major component of a person's identity, a simple prime in this study was enough to induce threat. Similarly, in another study, Asian-American women answered questions regarding either their gender or ethnicity before taking a math test (Shih, Pittinsky, & Ambady, 1999). Compared to a control group, these women performed better on a math test when their ethnic identity was activated, but worse when their gender identity was activated. As a final example, math performance of young girls (ages 5-7) whose mothers disagreed with the typical gender stereotype was unaffected by threat (Tomasetto, Alparone, & Cadinu, 2011). Girls whose mothers did not disagree with the stereotype, however, demonstrated poorer math performance when under threat in comparison to their pre-threat scores and the scores of the aforementioned girls. Altogether, these studies suggest that the stereotype threat can directly impact cognitive performance, but perhaps only when the stereotype is especially salient for an individual.

Studies examining the moderating role of group identification on threatened older adults' memory performance have reported findings consistent with those from other stigmatized groups. Older adults who strongly identified with their age group recalled fewer details from a prose passage than those who weakly identified (Kang & Chasteen, 2009). Hess and colleagues (Hess & Hinson, 2006; Hess et al., 2009) found that stereotype threat had the greatest negative impact on memory for older adults who were classified as young-old (i.e., 60-70) and/or were highly educated. They hypothesized that these findings may have been attributed to the fact that (1) young-old adults had just entered "older adulthood" and were thus more sensitive to stereotype labeling, and (2) highly educated older adults placed a greater value on their cognitive abilities and therefore were more sensitive to stereotyping.

Thomas and Dubois' older adult sample was younger than Wong and Gallo's (69.8 vs. 74.6 years old). In light of the Hess et al. (2009) findings, it is possible that a majority of older adults in the Thomas and Dubois (2011) experiment internalized the threat manipulation, while a majority of those in the Wong and Gallo (2015) study did not. Other individual differences such as level of education (Hess et al., 2009) and age-group identification (Kang & Chasteen, 2009) may have also contributed to older adults either being affected or unaffected by the stereotype passage. These individual differences may explain why only Thomas and Dubois (2011) observed increased false recognition for older adults under threat. That is, the discrepancy between Thomas and Dubois (2011) and Wong and Gallo (2015) could be due to slight methodological differences (i.e., the DRM warning) as was previously hypothesized (see Wong & Gallo, 2015), or differences in the extent to which their two older adult samples actually internalized the stereotype threat manipulation. Since the internalization of threat is contingent on individuals identifying with the stigmatized group, it is crucial to consider how measures of group identification may moderate the effects of threat on memory. These influences were not considered in the previous two experiments, but were addressed in the present study.

The Present Study

We aimed to determine the source of the inconsistency between Thomas and Dubois (2011) and Wong and Gallo (2015). We examined false and veridical memory performance for young and older adults in each of the four groups examined in those studies: negative stereotype threat, the DRM warning, threat and warning, or control (no passage). We also collected identity-related measures to determine whether group identification played a role in moderating the effects of stereotype threat on false memory susceptibility. Consistent with the previous two studies, we hypothesized that threatened older adults would demonstrate greater false recognition than the control group, and that those in the threat + warning group would demonstrate lower false recognition than the warning group. We further predicted that group identification measures would moderate the relationship between threat and false recognition, such that the patterns observed by Thomas and Dubois (2011) and Wong and Gallo (2015) would be most pronounced for individuals who were most likely to strongly identify with the stigmatized group.

Method

Participants

Three hundred adults participated in the experiment. Young adults (n = 100) were undergraduate students recruited from introductory psychology courses at Tufts University who received course credit (*M* age = 18.26, *SD* = 2.67, range = 17-23). Older adults (n = 200) were selected from a pre-established participation pool maintained by the Cognitive Aging and Memory Laboratory at Tufts University and were paid $15.00 for their participation (*M* age = 71.06, *SD* = 7.11, range = 56-93). To be a member of the participant pool, older adults were prescreened for psychological and neurological health issues, and could not be taking medications that might affect cognition. Young and older adults were tested separately; young adults were tested in groups of four and older adults were tested individually. Twenty-five young and 50 older adults were randomly assigned to the threat-only group, 25 young and 50 older adults to the warning-only group, 25 young and 50 older adults to the combined threat and warning group, and 25 young and 50 older adults to the control group. All participants completed the Vocabulary Subtest of the Shipley Institute of Living Scale 2 (see Appendix C; Shipley, 1946), on which older adults outperformed young adults, t (219.92) = 4.61, $p < .001$. Older adults also completed more years of formal education than young adults, t (296.99) = 11.05, $p <$ .001. All participants provided informed consent.

Materials

Stimuli. Young adult testing materials consisted of 36 word-lists, each of which contained the 15 strongest semantic associates of a related but non-presented lure word (Stadler, Roediger, & McDermott, 1999). Of these lists, 18 were studied and words from the other 18 lists served as distractors on the recognition test. Twenty-four lists (12 studied and 12 distractor) were used in older adult testing. Fewer lists were used with older adults to reduce the occurrence of task fatigue. We counterbalanced which lists served as stimuli and which served as lists from which we chose distractor items.

Recognition test. The recognition test for young adults consisted of 180 words: 18 words were the non-presented lures associated with the studied lists, 72 words had been studied, and 90 words were taken from non-presented lists. The 72 studied words on the recognition test consisted of the first, fifth, eighth, and tenth strongest associates of each related lure word (Stadler, Roediger, & McDermott, 1999). The 90 distractor items on the recognition test consisted of the related lure words from each of the 18 non-studied lists, as well as the first, fifth, eight, and tenth strongest associates from each of those lists. The older adult recognition test was constructed similarly, but contained 120 items due to the reduced number of lists that were studied. Stimuli and methodology were taken from standards in the DRM literature (see Gallo, 2006 for a review).

Threat and warning passages. The threat passage (209 words) described research that has demonstrated age-related declines in memory (see Appendix A; borrowed from Thomas & Dubois, 2011). It also stated that the current experiment was about memory and that the memory performance of older adults would be compared to that of young adults. The warning passage (176 words) explicitly revealed the deceptive nature of the DRM paradigm by drawing participants' attention to the fact that the words they had rated for pleasantness were associates of oft-falsely-remembered theme words (see Appendix B; borrowed from Gallo et al., 2001). The warning included an exemplar list (e.g., *marriage, vows, bride*) with a hypothetical non-presented lure (e.g., *wedding*), as well as instructions encouraging participants to avoid falsely recognizing related lures.

Identity questionnaire. The identity questionnaire consisted of questions designed to measure older adults' age-group identification (see Appendix D). This questionnaire included an adapted version of Pinel's (1999) Stigma Consciousness Questionnaire (SCQ) to examine the prevalence of stereotypes in older adults' lives (e.g., "I never worry that my behaviors will be viewed as stereotypically elderly"). Scores for the ten items were added together to create one composite score for each participant. Lower composite scores on the SCQ were indicative of greater concerns about age-related stigmas. Scores on this scale showed adequate internal consistency reliability (Cronbach's $\alpha = .77$).

Procedure

The experiment employed a 3 (age: young, young-old, or old-old) by 4 (passage type: threat, warning, threat and warning, or no passage) between-subjects factorial design. After providing informed consent, participants provided demographic information regarding age, gender, ethnicity, and total completed years of education since kindergarten. Participants then began the study phase. This phase consisted of an incidental-learning task, in which participants were presented with DRM wordlists. Lists were presented in blocked form at a rate of 1.5 s per word. List order and item order within each list were random. Participants were instructed to rate each word for pleasantness on an 8-point Likert-type scale (0 = not pleasant, 7 = extremely pleasant). After encoding, participants worked on a filler task for 5 min. The filler task consisted of a SUDOKU puzzle for young adults, and a 5 min break for older adults during which they talked casually with the experimenter. Older adults were not required to complete a SUDOKU puzzle because pilot testing revealed that many older adults found the task to be confusing and strenuous.

Participants were then randomly assigned to read the threat passage, warning passage, both passages, or no passage (control). Reading of passages was self-paced. Young and older participants read the same stereotype threat and warning passages. Following, all participants completed a yes/no recognition test, in which they indicated whether they had previously rated each word for pleasantness. Because they did not read any passages, participants in the control group had a slightly shorter retention interval between encoding and the recognition test. After each recognition judgment, participants rated their confidence in that judgment on an 8-point Likert-type scale (0 = not confident, 7 = extremely confident). All participants then took a vocabulary test (Shipley, 1946) and were debriefed. The identity questionnaire was administered to 160 of the 200 older adults who were tested. Eighty-two of these older adults completed the identity questionnaire prior to debriefing, and 78 older adults completed the questionnaire over the phone at a later date. Young adults did not complete the identity questionnaire because the questions were only relevant to older adults (e.g., “Are you retired?”).

The initial encoding, threat and warning instructions, and yes/no recognition test were presented using E-Prime software (Version 2.1; Schneider, Eschman, & Zuccolotto, 2001). Additionally, it should be noted that young adults completed the computer-based aspects of the experiment on their own. However, to minimize differences in computer experience, the experimenter entered all computer-based responses during older adult testing. The experiment took approximately 45 min.

Results and Discussion

Memory Accuracy

Hits. We conducted a 3 (age: young, young-old, and old-old) x 4 (group: threat, warning, threat + warning, control) between-subjects analysis of variance (ANOVA) on average hit proportions. Hit proportions were calculated by dividing the number of correctly recognized studied items by the total number of studied items that occurred on the recognition test for each participant. We found a main effect of age, F (2, 288) = 15.72, $p < .001$, $\eta_p^2 = .098$. Post-hoc tests using the Bonferroni correction showed that young adults outperformed young-old (.90 vs. .84; t (184.51) = 4.61, $p < .001$, $d = .65$) and old-old adults (.90 vs. .82; t (164.32) = 5.29, $p < .001$, $d = .75$), but young-old and old-old adults' performance did not differ. We did not observe a main effect of group or an interaction. See Table 1 for average hit proportions.

For the purposes of directly replicating Thomas and Dubois (2011), we ran a 2 (age: young adult vs. older adult) x 2 (group: threat vs. control) ANOVA on hit proportions. All older adults were categorized together in this analysis. Consistent with their findings, we found a main effect of age as young adults outperformed older adults (.88 vs. .83), F (1, 146) = 8.67, $p = .004$, $\eta_p^2 = .056$. There were no other significant main effects. The interaction between the two variables was also non-significant.

To more closely examine the differences in memory performance that were previously observed for young-old versus old-old adults (Hess et al., 2009), we also conducted a 2 (age: young-old vs. old-old) x 2 (group: threat vs. control) ANOVA on hit proportions. We again found a main effect of age, showing that young-old adults outperformed old-old adults (.85 vs.

.81), F (1, 96) = 5.07, p = .027, η_p^2 = .050, but no main effect of group or an interaction.

We also conducted a 2 (age: young adult vs. older adult) x 2 (group: warning vs. threat + warning) between-subjects ANOVA on hit proportions, to determine whether our data were consistent with findings reported in Wong and Gallo (2015). We observed a main effect of age as young adults had higher hit proportions than older adults (.91 vs. .83), F (1, 146) = 21.71, p < .001, η_p^2 = .129, but no main effect of group. Further, we did not observe the age by group interaction that Wong and Gallo (2015) found.

In another comparison to Wong and Gallo (2015), we examined age-related differences within the older adult population via a 2 (age: young-old vs. old-old) x 2 (group: warning vs. threat + warning) ANOVA on hit proportions. This analysis did not yield any significant main effects or an interaction.

In sum, our analyses on hit proportions revealed typical age-related memory differences as young adults accurately remembered more previously studied words than older adults. Furthermore, within the threat and control groups, young-old adults demonstrated better memory performance than old-old adults, as measured by hits. This finding is also consistent with previous aging literature showing that memory accuracy in the DRM paradigm declines as a function of age (Balota et al., 1999). Finally, our analyses on hit proportions replicated Thomas and Dubois (2011) but did not replicate Wong and Gallo (2015).

False Alarms. False alarm proportions were calculated by dividing the number of related lure words that each participant falsely recognized by the total number of related lure words presented on the recognition test. A 3 (age) x 4 (group) between-subjects ANOVA on average false alarm proportions revealed a main effect of age, F (2, 288) = 4.48, p = .012, η_p^2 = .030. Post-hoc tests using the Bonferroni correction showed that young adults falsely recognized fewer related lure words than young-old adults (.56 vs. .65; *t* (198) = 2.92, *p* = .010, *d* = .41), but there were no differences between young and old-old adults' performance (.56 vs. .59, *t* (198) = 1.10, *p* = *NS*) or young-old and old-old adults' performance (.65 vs. .59, *t* (198) = 1.86, *p* = *NS*). We did not observe a significant main effect of group. The interaction between the two variables was also non-significant. See Table 1 for average false alarm proportions.

As with hit proportions, we conducted 2 x 2 ANOVAs to directly compare false alarm proportions to those of Thomas and Dubois (2011) and Wong and Gallo (2015). Our 2 (age: young adult vs. older adult) x 2 (group: threat vs. control) between-subjects ANOVA on false alarm proportions yielded neither significant main effects nor an interaction (all *p*'s > .40), and thus failed to replicated Thomas and Dubois (2011). As was done with hit proportions, we also examined false alarm proportions only for the older adult sample. This 2 (age: young-old vs. old-old) by 2 (group: threat vs. control) between-subjects ANOVA did not reveal any significant main effects or an interaction.

A 2 (age: young adult vs. older adult) x 2 (group: warning vs. threat + warning) ANOVA on false alarm proportions also did not reveal a significant interaction (*p* > .80), and thus failed to replicate Wong and Gallo (2015). We did find a main effect of age, with young adults recognizing fewer related lures than older adults (.49 vs. .62), *F* (1, 146) = 10.38, *p* = .002, η_p^2 = .066, but no main effect of group. We again compared performance to Wong and Gallo (2015) within the older adult sample via a 2 (age: young-old vs. old-old) x 2 (group: threat vs. control) ANOVA on false alarm proportions. We did not find any significant main effects or an interaction.

In summary, the analyses comparing false alarm proportions to those of Thomas and Dubois (2011) and Wong and Gallo (2015) failed to replicate either study. Additionally, we found evidence that false recognition in the DRM paradigm may not always increase as a function of age. This inconsistency is addressed in the subsequent regression analysis.

Corrected False Alarms. We next conducted similar ANOVAs on the high-threshold false alarm correction suggested by Gallo et al. (2001). The high-threshold correction was calculated by subtracting the proportion of false alarms to non-studied lure words (i.e., the lure words taken from distractor lists) from the proportion of false alarms to related lures. This technique serves to roughly eliminate the influence of unknown response biases (e.g., a participant's response criterion) on false alarm rates.

In accordance with the analyses on false alarm proportions, the 3 (age) x 4 (group) between-subjects ANOVA on corrected false alarms revealed a main effect of age, F (2, 288) = 8.76, $p < .001$, $\eta_p^2 = .057$, as young adults demonstrated lower false recognition than young-old adults (.39 vs. .51; t (198) = 4.06, $p < .001$, $d = .57$) but not old-old adults (.39 vs. .45; t (198) = 2.04, $p = NS$). Again, this provides evidence that false memory susceptibility may not always increase with age in older adulthood. See Table 1 for corrected false alarm averages.

The 2 (age: young adult vs. older adult) x 2 (group: threat vs. control) ANOVA comparing our data to those of Thomas and Dubois (2011) found no significant main effects or an interaction (all *p*'s > .20). When focusing solely on the older adult sample, a 2 (age: young-old vs. old-old) by 2 (group: threat vs. control) ANOVA on corrected false alarms also did not reveal any significant effects (all *p*'s > .25).

The 2 (age: young adult vs. older adult) x 2 (group: warning vs. threat + warning) ANOVA that served as the Wong and Gallo (2015) comparison revealed a main effect of age as young adults once again demonstrated lower false recognition than older adults (.33 vs. 48), F (1, 146) = 15.83, $p < .001$, $\eta_p^2 = .098$. We did not find a significant main effect of group or an interaction between age and group.

In another comparison to Wong and Gallo (2015), we conducted a 2 (age: young-old vs. old-old) by 2 (group: warning vs. threat + warning) ANOVA on corrected false alarms. We found a main effect of age, F (1, 96) = 5.01, p = .027, η_p^2 = .050, as old-old adults demonstrated lower false recognition than young-old adults (.43 vs. .52). We also found a main effect of group, F (1, 96) = 6.15, p = .015, η_p^2 = .060, with older adults in the warning group demonstrating lower false recognition than those in the threat + warning group (.42 vs. .53).

Similar to the analyses on false alarm proportions, the ANOVAs conducted on the high-threshold false alarm correction did not provide support for either Thomas and Dubois (2011) or Wong and Gallo (2015). False recognition was lower for older adults in the warning group than in the threat + warning group, contradicting the Wong and Gallo (2015) findings.

Zhang and Mueller's A. We last examined memory accuracy via Zhang and Mueller's (2005) non-parametric A statistic. The A statistic is an average of the maximum-area and minimum-area proper ROC curves through each observed point, where each point is a participant's average proportion of hits and proportion of false alarms to related lures plotted in a Cartesian plane. Higher A values are indicative of relatively high hit rates and relatively low false alarm rates, providing a measure of "good" memory performance. In the context of the DRM paradigm, good memory performance involves maximizing hits and reducing false alarms by carefully discriminating between previously studied words and related lures at test.

A 3 (age) x 4 (group) ANOVA on A scores revealed a main effect of age, F (2, 288) = 9.44, p < .001, η_p^2 = .062. Post-hoc tests using the Bonferroni correction showed that young adults had higher A scores than young-old adults (.73 vs. .57; t (174.42) = 4.26, p < .001, d = .60) and old-old adults (.73 vs. .63; t (198) = 3.11, p = .016, d = .44), but there was no difference in A scores between young-old and old-old adults. That is, young adults were better able to discriminate between previously studied and not studied material than all older adults. We did not observe a main effect of group or an interaction. See Table 1 for average A scores.

A 2 (age: young adult vs. older adult) x 2 (group: threat vs. control) ANOVA on A scores found a main effect of age as young adults had higher A scores than older adults (.70 vs. .61), F (1, 146) = 3.96, $p = .048$, $\eta_p^2 = .026$. When examining performance for older adults only, a 2 (age: young-old vs. old-old) by 2 (group: threat vs. control) ANOVA on A scores revealed a significant interaction, F (1, 96) = 4.69, $p = .033$, $\eta_p^2 = .047$ (see Figure 1). Independent samples t tests showed that young-old adults' A scores did not differ between the threat and control groups (.61 vs. .55; t (48) = 0.61, $p = .544$) whereas old-old adults had significantly lower A scores in the threat group than in the control group (.56 vs. .73; t (31.83) = 3.06, $p = .004$, d = .87). This finding provides support for Thomas and Dubois (2011), but only when the population of interest is adults ages 71 and older. It appears that stereotype threat exerted subtle effects on old-old adults' hit and false alarm proportions that were not detectable until the two dependent measures were examined together via the A statistic.

Finally, the 2 (age: young adult vs. older adult) x 2 (group: warning vs. threat + warning) ANOVA revealed a main effect of age as young adults outperformed older adults (.77 vs. .59), F (1, 146) = 13.45, $p < .001$, $\eta_p^2 = .084$. The main effect of group and the interaction were non-significant. The 2 (age: young-old vs. old-old) x 2 (group: warning vs. threat + warning) ANOVA focusing solely on the older adult sample did not find any significant main effects or an interaction.

Analyses on Zhang and Mueller's (2005) A statistic replicated our previous analyses as young adults had higher scores than older adults, demonstrating their superior ability to discriminate between studied and non-studied items at test. We also provided support for Thomas and Dubois (2011) as old-old adults, but not young-old adults, had significantly lower A scores in the threat group than in the control group. We did not find any evidence to support Wong and Gallo (2015).

Meta-cognitive Judgments

Average Confidence. A 3 (age: young, young-old, or old-old) x 4 (group: threat, warning, threat + warning, or control) between-subjects ANOVA was conducted on average confidence ratings associated with false alarms to related lures. All effects were non-significant (all *p*'s > .10). See Table 3 for average confidence ratings for hit and false alarm proportions.

Gamma Correlations. We computed Goodman-Kruskal gamma correlations for each participant. Gamma correlations are an analysis of the relationship between meta-cognitive predictions or post-dictions and criterion performance (Nelson, 1984, p. 109). Higher gamma values are indicative of better meta-cognitive accuracy. We examined post-diction accuracy by correlating performance on each recognition test item with each subsequent confidence rating.

Separate gamma correlations were computed for previously studied recognition test items and related lure words (see Table 4 for averages). Because gamma correlations cannot be computed for participants who provide the same confidence rating for every answer on the recognition test, some participants were excluded from the analysis. This left 271 participants (88 young, 91 young-old, 92 old-old) in the analyses on studied items and 268 participants (92 young, 87 young-old, 89 old-old) in the analyses on related lure words.

We conducted 3 (age: young, young-old, or old-old) x 4 (group: threat, warning, threat + warning, or control) ANOVAs on gamma correlations for studied items and related lure words. The ANOVA on studied items yielded a main effect of age, F (2, 259) = 11.19, $p < .001$, η_p^2 =

.080. Bonferroni post-hoc comparisons showed that young adults had higher gamma correlations than young-old (.88 vs. .74; t (120.99) = 3.61, p = .027, d = .54) and old-old adults (.88 vs. .62; t (107.11) = 4.87, p < .001, d = .72), and young-old adults had a marginally significantly higher average gamma than old-old adults (.74 vs. .62; t (163.55) = 1.89, p = .081, d = .23).

Multiple Regression

In the two previous studies that examined older adults' false memory susceptibility when under stereotype threat, researchers focused on age as the sole predictor of memory performance (Thomas & Dubois, 2011; Wong & Gallo, 2015). The results from these experiments were inconsistent, and our ANOVA comparisons failed to directly replicate either of them. Furthermore, when we divided our older adult age group into young-old and old-old adults, the results of our analyses on A scores also failed to replicate Hess et al. (2009). All of these discrepancies suggest that age alone may not be a reliable predictor of older adults' false memory susceptibility when under stereotype threat.

In a small body of research investigating the effects of threat on older adults' veridical memory performance, researchers took a multivariate approach. They showed that, in addition to age, variables such as years of education, level of sensitivity to aging stereotypes, and level of identification with the older adult age group influenced older adults' memory accuracy when under stereotype threat (Hess et al., 2009; Kang & Chasteen, 2009). These findings suggest that components of one's identity may help predict threat-related false memory susceptibility. To test this hypothesis, we used multiple linear regression to model the contributions of stereotype threat, age, years of education, and group-identification measures to variability in older adults' false recognition performance. Because we were specifically interested in identifying moderators of the effects of stereotype threat on false recognition, only the threat and control groups were contrasted in these analyses.

Recall that 160 older adults completed the identity questionnaire. Forty-three of these older adults were in the threat group, and 34 older adults were in the control group. Thus, the following regression analyses were conducted using this reduced sample of 77 older adults (see Table 5 for frequencies).

To determine which of our predictors of interest were contributing to variability in our data, we first conducted an exploratory multiple linear regression on false alarm proportions for related lures. We examined main effects of group (threat vs. control), age, years of education, retirement status (yes/no), and SCQ score, as well as all possible two-way interactions among these five variables (see Table 6 for the correlation matrix for continuous variables). This model yielded a non-significant R^2 of .27 (F (15, 61) = 1.47, p = .146). We found a marginally significant main effect of group (β = .489, t = 1.79, p = .079), a significant group by retirement status interaction (β = .692, t = 2.20, p = .031), and a significant retirement status by years of education interaction (β = -.626, t = 2.02, p = .048). See Table 7 for regression coefficients for this model.

Because age, SCQ score, and their interactions with other variables did not contribute to a significant or marginally significant amount of variability in our initial regression, we eliminated them from the final model. In this model, we examined main effects of group, years of education, retirement status, all two-way interactions between these variables, and the three-way interaction. This analysis yielded a significant R^2 of .19, F (7, 69) = 2.29, p = .037, with three significant effects (see Table 8 for regression coefficients). Most notably, we found a main effect of group (β = .547, t = 2.26, p = .027) as older adults under stereotype threat falsely recognized more lure words than those in the control group. Thus, when years of education and retirement status were held constant, we provided a strong replication of Thomas and Dubois (2011).

Our analysis also revealed a group by retirement status interaction (see Figure 2; β = .727, $t = 2.65$, $p = .010$). In comparison to working older adults, retired older adults had marginally higher false alarm proportions when under stereotype threat ($t\,(41) = 1.41$, $p = .084$), but lower false alarm proportions in the control group ($t\,(32) = 1.84$, $p = .038$). Thus, retired older adults under stereotype threat were more susceptible to false memories than those who were still working.

Our regression on false alarm proportions also found a group by years of education interaction ($\beta = .989$, $t = 2.21$, $p = .030$). As shown in Figure 3, simple slopes analysis revealed that the interaction was due to a marginally significant decrease in false alarm proportions as years of education decreased for older adults in the control group ($b = -.023$, $t = 1.85$, $p = .069$), with no change in performance as a function of education for those in the threat group ($b = -.009$, $t = 0.93$, $p = .354$). That is, highly educated older adults in the control group were able to outperform those who were less educated, but this did not hold true for older adults in the threat group.

In sum, our multiple regression analysis on false alarm proportions for older adults in the threat and control groups replicated the Thomas and Dubois (2011) results, and revealed two key predictors of performance. Retired older adults demonstrated higher false recognition than still-working older adults when under threat, but lower false recognition in the control group. Additionally, while higher levels of education were associated with lower false recognition in the control group, these benefits disappeared in the presence of stereotype threat. Based on these findings, we would predict that stereotype threat would be most likely to increase false memory susceptibility for older adults who are retired and/or are highly educated.

Revisiting Thomas and Dubois (2011) and Wong and Gallo (2015)

We identified two variables that strongly moderated the effects of threat on false recognition of related lure words: retirement status and years of education. As such, we re-examined the Thomas and Dubois (2011) and Wong and Gallo (2015) comparisons, but only for older adults who were retired and/or highly educated. We defined "highly educated" as the mean years of education for older adults (M = 16.27) plus one standard deviation (SD = 2.98). Rounded down, this was 19 years of education. We used this method to determine what constituted "high" years of education because it was the same method used in simple slopes analysis to examine "high" values of a continuous variable. After eliminating from the sample older adults who did not meet these criteria, our remaining sample consisted of 129 older adults: 36 in the threat group, 35 in the warning group, 35 in the threat + warning group, and 23 in the control group.

An independent samples t test on this reduced sample showed that older adults under threat falsely recognized more related lures than those in the control group (.68 vs. .58; t (57) = 1.83, p = .036, d = .50), thus replicating Thomas and Dubois (2011). However, we again failed to replicate Wong and Gallo (2015) as older adults in the warning group did not perform significantly differently from those in the threat + warning group (.63 vs. .68; t (68) = 1.13, p = .132). In fact, the means were in opposite direction of what we had predicted.

General Discussion

We had two main goals in the present experiment: (1) to examine whether individual differences related to group identification moderate the effects of stereotype threat on false memory susceptibility, and (2) to determine whether these influences could explain why some researchers have found increases in false recognition for older adults under threat (Thomas & Dubois, 2011) and some have reported decreases (Wong & Gallo, 2015). We found that the effects of threat on false recognition were moderated by retirement status and years of education, such that threat resulted in the greatest false memory increases for older adults who were retired and/or highly educated. Older adults under stereotype threat who met either or both of those criteria demonstrated increased false recognition in comparison to our control group, replicating previous work (i.e., Thomas & Dubois, 2011). We did not, however, provide support for the finding that stereotype threat combined with a DRM warning resulted in lower false recognition than a warning-only group (i.e., Wong & Gallo, 2015).

The decreased false memory susceptibility that Wong and Gallo (2015) observed for older adults in their stereotype threat and warning group was interpreted within the regulatory fit framework. According to the regulatory fit hypothesis, stereotype threat induces a risk-averse mindset. When this mindset is matched with a risk-averse memory test, threatened individuals experience a “fit” between their attitude and the task that they have to perform, which results in superior memory performance. Because the DRM warning encourages risk-aversive responding on the recognition test, Wong and Gallo (2015) speculated that a regulatory fit was achieved for those who received both the threat and warning manipulations. Not only did we not replicate this pattern in the present study, we found that older adults in the warning group actually demonstrated *lower* false recognition than those in the threat + warning group.

Wong and Gallo’s (2015) regulatory fit explanation was based on the premise that stereotype threat had been effectively induced in their older adult participants. As such, we re-tested the Wong and Gallo (2015) hypothesis, limiting our sample to older adults who demonstrated the greatest susceptibility to stereotype threat. We identified those who were most susceptible to threat by comparing older adults' performance in the control and threat groups, and determining which older adults showed the greatest increases in false recognition in the threat group (i.e., those who were retired and/or highly educated). Upon re-analyzing, we still found no evidence of a regulatory fit occurring in the threat + warning group. This suggests that stereotype threat may not have actually been induced in Wong and Gallo's (2015) participants, discrediting a regulatory fit interpretation of their findings.

We propose an alternative explanation for the Wong and Gallo (2015) findings. Stereotype threat manipulations can be ineffective with individuals who do not identify with their stigmatized group, i.e., outgroup members (e.g., Tomasetto et al., 2011). When faced with a negative stereotype about members of a group with which they do not identify, outgroup members may experience *stereotype lift*. This is an improvement in performance that is observed when downward comparisons are made with a stigmatized outgroup (Walton & Cohen, 2003). If, as our analyses suggested, Wong and Gallo's (2015) older adults did not identify with the stereotype threat manipulation, they may have experienced this stereotype lift. The performance boost associated with stereotype lift accounts nicely for the relatively low false recognition rates for older adults under threat in the Wong and Gallo (2015) study.

In addition to Wong and Gallo (2015), several studies have interpreted threat-related memory improvements in terms of a regulatory fit (Barber & Mather, 2013; Chalabaev et al., 2012; Grimm et al., 2009; Maddox et al., 2006). Could these performance boosts also be reinterpreted as stereotype lift? That question could be answered via two different approaches. First, using a similar approach as the present study, researchers could identify individuals who experience typical threat-related memory deficits under normal circumstances in which a regulatory fit would not be present. Previous experiments could then be replicated using a sample of people who possess similar characteristics to those that experienced stereotype threat effects. Alternatively, these previous studies could be replicated and expanded upon to include measures related to identification with the stigmatized group being targeted. Results would then need to be analyzed only for participants who possess characteristics that make them highly likely to identify with the stereotyped group. Testing the regulatory fit hypothesis with the proposed group-identification considerations is an important direction for future research.

Identifying with the Stigmatized Group

The discrepancy in findings between Thomas and Dubois (2011) and Wong and Gallo (2015) can thus be interpreted in terms of older adult identity. We hypothesize that older adult participants in the present study and those in the Thomas and Dubois (2011) experiment possessed characteristics that made them more likely to identify with the groups that were stigmatized in the threat manipulation. This increased susceptibility to threat resulted in the threat-related memory deficits that we and Thomas and Dubois (2011) observed. One explanation for the conflicting findings presented by Wong and Gallo (2015) is that their participants may not have identified with the groups targeted by the stereotype threat manipulation. Failure to identify could result in a stereotype lift for those in the threat group. Future researchers should consider testing this hypothesis by examining false memory performance for older adults in a threat + warning group who are *not* predicted to identify with the stereotypes presented in the threat manipulation. Older adults who are less susceptible to stereotype threat, and thus more susceptible to stereotype lift, may include those who are not retired and those with relatively low levels of education.

How do retirement status and years of education help explain why some older adults identify with the threat manipulation and others do not? Individuals who strongly identify with their stereotyped group have been shown to experience stereotype-based decreases in cognitive performance (e.g., Shih et al., 1999; Tomasetto et al., 2011). Our stereotype threat manipulation emphasized poor cognitive performance for people over the age of 55. As such, the passage targeted those who identify as old and who value their intellectual abilities. Retirement is accompanied by a host of lifestyle changes that may result in older adults' identifying more strongly with their "older" age group. Additionally, individuals who are more highly educated may place more value on their cognitive abilities, thus increasing the likelihood that they would internalize threats to their intellect. In sum, stereotypes about memory and aging may be especially salient for retired and/or highly educated older adults, resulting in the observed increases in false recognition for threatened individuals who met these criteria. Our study is the second to identify a relationship between high levels of education and poor memory performance for older adults under threat (see Hess et al., 2009), and the first to demonstrate increased false memory susceptibility for threatened individuals who are retired.

Finally, with regard to theories of the mechanisms underlying stereotype threat effect, our experiment was not designed to directly test either the executive resource depletion or regulatory fit hypotheses. However, our findings continue to support a resource depletion explanation, but do not provide evidence of a regulatory fit. We replicated the finding that stereotype threat increases false memory susceptibility, which could be the result of a threat-induced resource depletion. Individuals who were affected by the stereotype threat passage may have been overwhelmed with threat-related thoughts, thus reducing the pool of cognitive resources available to them for engaging in careful recollection at test. Had a regulatory fit been achieved, older adults would have demonstrated increased hit rates or decreased false alarm rates when stereotype threat was combined with a DRM warning that encouraged risk-aversion (Barber & Mather, 2014; Wong & Gallo, 2015). However, neither pattern of responding was demonstrated.

Adding to the wealth of literature showing stereotype-threat-related performance decrements, we showed that *some* older adults experience increased false memory susceptibility when confronted with ageist stereotypes. We emphasize *some* because whether stereotype threat negatively affected memory was contingent upon retirement status and/or level of education. Certainly these individual differences are not the only measures that predict older adults' susceptibility to stereotype threat and, as a result, susceptibility to false memories. A clear limitation to this study was the small amount of demographic information collected, and the few measures of age-group identification taken. Future researchers should expand the search for measures of identity that help predict which older adults are most likely to internalize negative aging stereotypes. Identifying the types of people who are at risk of stereotype threat effects is an important step toward determining ways to help older adults reduce their false memory susceptibility.

References

Balota, D. A., Cortese, M. J., Duchek, J. M., Adams, D., Roediger, H. L., McDermott, K. B., & Yerys, B. E. (1999). Veridical and false memories in healthy older adults and in dementia of the alzheimer's type. *Cognitive Neuropsychology,16*(3-5), 361-384.

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

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General, 136*(2), 256-276. doi:http://dx.doi.org/10.1037/0096-3445.136.2.256

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science, 11*(5), 164-169. doi:http://dx.doi.org/10.1111/1467-8721.00192

Brainerd, C. J., & Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review, 10*(1), 3-47. Retrieved from http://search.proquest.com.ezproxy.library.tufts.edu/docview/617747303?accountid=14434

Chalabaev, A., Major, B., Sarrazin, P., & Cury, F. (2012). When avoiding failure improves performance: Stereotype threat and the impact of performance goals. *Motivation and Emotion, 36*(2), 130-142.

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*(3), 235-260. doi:http://dx.doi.org/10.1080/03610730590948177

Dehon, H., & Brédart, S. (2004). False memories: Young and older adults think of semantic associates at the same rate, but young adults are more successful at source monitoring. *Psychology and Aging, 19*(1), 191-197. doi:http://dx.doi.org/10.1037/0882-7974.19.1.191

Gallo, D. A. (2006). *Associative illusions of memory: False memory research in DRM and related tasks* Psychology Press, New York, NY. Retrieved from http://search.proquest.com.ezproxy.library.tufts.edu/docview/621479228?accountid=14434

Gallo, D. A., Roediger, H. L., & McDermott, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin & Review, 8*(3), 579-586. doi:http://dx.doi.org/10.3758/BF03196194

Grimm, L. R., Markman, A. B., Maddox, W. T., & Baldwin, G. C. (2009). Stereotype threat reinterpreted as a regulatory mismatch. *Journal of Personality and Social Psychology, 96*(2), 288-304. doi:http://dx.doi.org/10.1037/a0013463

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, 58B*(1), P3-P11. doi:http://dx.doi.org/10.1093/geronb/58.1.P3

Hess, T. M., Emery, L., & Queen, T. L. (2009). Task demands moderate stereotype threat effects on memory performance. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 64B*(4), 482-486. doi:http://dx.doi.org/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*(3), 621-625. doi:http://dx.doi.org/10.1037/0882-7974.21.3.621

Hess, T. M., Hinson, J. T., & Hodges, E. A. (2009). Moderators of and mechanisms underlying stereotype threat effects on older adults' memory performance. *Experimental Aging Research, 35*(2), 153-177. doi:http://dx.doi.org/10.1080/03610730802716413

Kang, S. K., & Chasteen, A. L. (2009). The moderating role of age-group identification and perceived threat on stereotype threat among older adults. *The International Journal of Aging & Human Development, 69*(3), 201-220. doi:http://dx.doi.org/10.2190/AG.69.3.c

Maddox, W. T., Baldwin, G. C., & Markman, A. B. (2006). A test of the regulatory fit hypothesis in perceptual classification learning. *Memory & Cognition, 34*(7), 1377-1397. doi:http://dx.doi.org/10.3758/BF03195904

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

McCabe, D. P., & Smith, A. D. (2002). The effect of warnings on false memories in young and older adults. *Memory & Cognition, 30*(7), 1065-1077. doi:http://dx.doi.org/10.3758/BF03194324

McDermott, K. B., & Watson, J. M. (2001). The rise and fall of false recall: The impact of presentation duration. *Journal of Memory and Language, 45*(1), 160-176. doi:http://dx.doi.org/10.1006/jmla.2000.2771

McIntyre, J. S., & Craik, F. I. (1987). Age differences in memory for item and source information. *Canadian Journal of Psychology/Revue Canadienne De Psychologie, 41*(2), 175-192. doi:http://dx.doi.org/10.1037/h0084154

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95,* 109-133.

Neuschatz, J. S., Payne, D. G., Lampinen, J. M., & Toglia, M. P. (2001). Assessing the effectiveness of warnings and the phenomenological characteristics of false memories. *Memory, 9*(1), 53-71. doi:http://dx.doi.org/10.1080/09658210042000076

Pinel, E. C. (1999). Stigma consciousness: The psychological legacy of social stereotypes. *Journal of Personality and Social Psychology, 76*(1), 114-128. doi:http://dx.doi.org/10.1037/0022-3514.76.1.114

Régner, I., Smeding, A., Gimmig, D., Thinus-Blanc, C., Monteil, J., & Huguet, P. (2010). Individual differences in working memory moderate stereotype-threat effects. *Psychological Science, 21*(11), 1646-1648. doi:http://dx.doi.org/10.1177/0956797610386619

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7*(1), 1-75. Retrieved from http://search.proquest.com.ezproxy.library.tufts.edu/docview/618691046?accountid=14434

Roediger, H. L., Balota, D. A., & Watson, J. M. (2001). Spreading activation and arousal of false memories. *The nature of remembering: Essays in honor of robert G. crowder*. (pp. 95-115) American Psychological Association, Washington, DC. doi:http://dx.doi.org/10.1037/10394-006

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803-814. doi:http://dx.doi.org/10.1037/0278-7393.21.4.803

Schmader, T. (2010). Stereotype threat deconstructed. *Current Directions in Psychological Science, 19*(1), 14-18. doi:http://dx.doi.org/10.1177/0963721409359292

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

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

Schneider, W., Eschman, A., & Zuccolotto, A. (2001). E-Prime user's guide. Pittsburgh, PA: Psychology Software Tools.

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

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*(1), 80-83. Retrieved from http://search.proquest.com.ezproxy.library.tufts.edu/docview/619416244?accountid=14434

Shipley, W. C. (1946). Institute of Living Scale. Los Angeles, CA: Western Psychological Services.

Stadler, M. A., Roediger, H. L., & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition, 27*(3), 494-500. Retrieved from http://search.proquest.com.ezproxy.library.tufts.edu/docview/619409432?accountid=14434

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

Stein, R., Blanchard-Fields, F., & Hertzog, C. (2002). The effects of age-stereotype priming on the memory performance of older adults. *Experimental Aging Research, 28*(2), 169-181. doi:http://dx.doi.org/10.1080/03610730252800184

Thomas, A. K., & Bulevich, J. B. (2006). Effective cue utilization reduces memory errors in older adults. *Psychology and Aging,21*(2), 379-389. doi:http://dx.doi.org/10.1037/0882-7974.21.2.379

Thomas, A. K., Bulevich, J. B., & Chan, J. C. K. (2010). Testing promotes eyewitness accuracy with a warning: Implications for retrieval enhanced suggestibility. *Journal of Memory and Language, 63*(2), 149-157. doi:http://dx.doi.org/10.1016/j.jml.2010.04.004

Thomas, A. K., & Dubois, S. J. (2011). Reducing the burden of stereotype threat eliminates age differences in memory distortion. *Psychological Science*, *22*(12), 1515-1517.

Thomas, A. K., & Sommers, M. S. (2005). Attention to item-specific processing eliminates age effects in false memories. *Journal of Memory and Language, 52*(1), 71-86. doi:http://dx.doi.org/10.1016/j.jml.2004.08.001

Tomasetto, C., Alparone, F. R., & Cadinu, M. (2011). Girls' math performance under stereotype threat: The moderating role of mothers' gender stereotypes. *Developmental Psychology, 47*(4), 943-949. doi:http://dx.doi.org/10.1037/a0024047

Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology and Aging, 13*(2), 230-241. doi:http://dx.doi.org/10.1037/0882-7974.13.2.230

Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology, 39*(5), 456-467. doi:http://dx.doi.org/10.1016/S0022-1031(03)00019-2

Wong, J. T., & Gallo, D. A. (in press). Stereotype threat reduces false recognition when older adults are forewarned. *Memory*.

Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika, 70*(1), 1-10. doi:http://dx.doi.org/10.1007/s11336-003-1119-8

Appendix A

Stereotype Threat Passage

It is a common belief that memory declines with age. Adults over age 55 report more memory problems in everyday life than do young adults, and older adults also report less perceived control over their own memories. Such beliefs are not groundless. Older adults perform worse than young adults on standardized memory tests, as well as on laboratory tests of both recall and recognition. Complaints regarding memory performance increase with age even for individuals in good health. Furthermore, age-related declines in memory cannot be attributed to artificiality of laboratory tasks, because older adults perform more poorly than young adults on tasks that are designed to simulate "real life."

For instance, it has been found that older adults have poorer memories for passages they read, for information on simulated medicine labels, and for activities they have performed. Their memory is poorer for names and faces of people, for theme songs of television programs they have watched, for the layout of museums they have visited—and the list goes on (Bjork & Bjork, 1996, p. 443-444).

In this experiment, we are interested in comparing how well older adults remember lists of words when they were not told to study the words, and their performances will be compared to college-aged young adults.

Appendix B

Warning Passage

Before we start the test, please be warned that some of the test items will be tricky. You may have noticed that each list of words for which you made pleasantness ratings was comprised of related words. In fact, all the words in each list were associated to one common word, and this common word was not actually presented in the list. Prior research indicates that people often falsely remember that this word was in the list. We want you to avoid making this memory error.

As an example, read the following list which was not used in this experiment:

marriage, ceremony, matrimony, bride, groom, vows, love, engagement, ring, honeymoon, dress, wife, husband, relationship, bridesmaid

In this case, wedding is the word that links all the above words. Even though it was not presented in the list, people often falsely remember it as having been on the list. That is, people can mistakenly remember the related word that links together all the others (e.g., wedding) even when it was not presented. Do not make this error.

Appendix C

VOCABULARY TEST

Participant #_______________ Date: _________________ Age_______ Sex: ______

Highest level of school completed (circle one): middle school some high school

high school some college college advanced degree

In each word in capital letters, underline the word in the group below it that has the same meaning. The first one with CONNECT is an example.

1. CONNECT
accident <u>join</u>
lace bean
flirt pierce

2. PROVIDE
harmonize commit
hurt supply
annoy divide

3. STUBBORN
obstinate steady
hopeful hollow
orderly slack

4. SCHOONER
building man
ship singer
plant scholar

5. LIBERTY
worry freedom
rich serviette
forest cheerful

6. COURTEOUS
dreadful proud
truthful short
curtsey polite

7. RESEMBLANCE
attendance fondness
assemble repose
likeness memory

8. THRIVE
flourish cry
thrash leap
field think

9. PRECISE
natural stupid
faulty grand
small exact

10. ELEVATE
revolve move
raise work
waver disperse

11. LAVISH
unaccountable selfish
romantic lawful
extravagant praise

12. SURMOUNT
mountain descend
overcome concede
appease snub

13. BOMBASTIC
democratic pompous
bickering cautious
destructive anxious

14. ENVISAGE
contemplate activate
surround estrange
enfeeble regress

15. PERPETUATE
appropriate commit
propitiate deface
blame control

16. LIBERTINE
missionary rescuer
profligate canard
regicide farrago

17. QUERULOUS
astringent fearful
petulant curious
inquiring spurious

18. FECUND
esculent optative
profound prolific
sublime salic

19. ABNEGATE
contradict decry
renounce execute
belie assemble

20. TRADUCE
challenge attenuate
suspend establish
misrepresent conclude

21. TEMERITY
impermanence rashness
nervousness stability
punctuality submissiveness

Appendix D

Identity Questionnaire

Subject number ____________

Date _____________

Are you retired? Please circle:

yes **yes but still work part-time** **no**

If you are retired, in what year did you retire? _______________

What is/was your occupation?

____

____

What extra-curricular activities do you engage in (for example, do you do any volunteer work? Do you participate in activities at churches, senior centers, or elsewhere?)

____

____

____

____

How often do you interact with adults between the ages of 18 and 30? Please circle the statement that best summarizes your interactions:

Several times per day

Once or twice per day

Several times per week

Once or twice per week

Several times per month

Once or twice per month

Several times per year

Once or twice per year

How would you describe your interactions with adults in that age group? Please circle:

Always pleasant

Usually pleasant

Sometimes pleasant

Neither pleasant nor unpleasant

Sometimes unpleasant

Usually unpleasant

Always unpleasant

Please rate the following statements on a scale of 0-6 to indicate how much you agree or disagree, where 0 = strongly disagree and 6 = strongly agree.

Please note that in the following statements, the term “older adult” refers to adults ages 60 and older, and the term “young adult” refers to adults ages 18-30.

0 = strongly disagree, 1 = moderately disagree, 2 = slightly disagree, 3 = neither agree nor disagree, 4 = slightly agree, 5 = moderately agree, 6 = strongly agree

1. Stereotypes about older adults have not affected me personally **0 1 2 3 4 5 6**
2. I never worry that my behaviors will be viewed as stereotypically elderly. **0 1 2 3 4 5 6**
3. When interacting with young adults, I feel like they interpret all my behaviors in terms of the fact that I am an older adult. **0 1 2 3 4 5 6**
4. Most young adults do not judge older adults on the basis of their age. **0 1 2 3 4 5 6**
5. My being an older adult does not influence how young adults act with me. **0 1 2 3 4 5 6**
6. I almost never think about the fact that I am older when I interact with young adults. **0 1 2 3 4 5 6**
7. Most young adults have a lot more ageist thoughts than they actually express. **0 1 2 3 4 5 6**
8. My being an older adult does not influence how people act with me. **0 1 2 3 4 5 6**
9. I often think that young adults are unfairly accused of being ageist. **0 1 2 3 4 5 6**
10. Most young adults have a problem viewing older adults as equals. **0 1 2 3 4 5 6**

Table 1

Average Proportions of Hits, Proportions of False Alarms to Related Lure Words, High-Threshold False Alarms, and A Scores.

Measure	Young Adults	Young-old Adults	Old-old Adults
Hits			
Threat	.88 (*.08*)	.85 (*.09*)	.77 (*.14*)
Warning	.92 (*.06*)	.83 (*.09*)	.81 (*.13*)
Threat and warning	.91 (*.08*)	.82 (*.13*)	.86 (*.11*)
Control	.88 (*.08*)	.86 (*.09*)	.84 (*.10*)
False Alarms to Related Lures			
Threat	.60 (*.24*)	.64 (*.26*)	.64 (*.20*)
Warning	.48 (*.24*)	.63 (*.18*)	.56 (*.23*)
Threat and warning	.50 (*.17*)	.66 (*.23*)	.61 (*.25*)
Control	.63 (*.18*)	.66 (*.25*)	.55 (*.17*)
High-Threshold False Alarms			
Threat	.42 (*.23*)	.50 (*.22*)	.50 (*.21*)
Warning	.33 (*.21*)	.47 (*.19*)	.38 (*.20*)
Threat and warning	.34 (*.17*)	.57 (*.22*)	.48 (*.22*)
Control	.46 (*.19*)	.49 (*.24*)	.42 (*.15*)
A Scores			
Threat	.71 (*.18*)	.61 (*.29*)	.56 (*.26*)
Warning	.79 (*.19*)	.58 (*.31*)	.64 (*.26*)
Threat and warning	.74 (*.24*)	.55 (*.32*)	.58 (*.35*)
Control	.68 (*.23*)	.55 (*.33*)	.73 (*.11*)

Note: Standard deviations are given in parentheses.

Table 2

Average Proportions of False Alarms to Nonstudied Lure Words and False Alarms to Nonstudied List Items

Measure	**Young Adults**	**Young-old Adults**	**Old-old Adults**
False Alarms to Nonstudied Lures			
Threat	.18 (*.13*)	.14 (*.13*)	.14 (*.10*)
Warning	.16 (*.11*)	.16 (*.20*)	.19 (*.14*)
Threat and warning	.17 (*.10*)	.08 (*.10*)	.13 (*.10*)
Control	.17 (*.14*)	.18 (*.17*)	.12 (*.11*)
False Alarms to Nonstudied List Items			
Threat	.13 (*.09*)	.09 (*.09*)	.09 (*.07*)
Warning	.12 (*.09*)	.12 (*.13*)	.12 (*.14*)
Threat and warning	.12 (*.07*)	.06 (*.05*)	.09 (*.07*)
Control	.13 (*.08*)	.12 (*.13*)	.07 (*.05*)

Note: Standard deviations are given in parentheses.

Table 3

Average Confidence Ratings for Hits and False Alarms to Related Lures.

Measure	**Young Adults**	**Young-old Adults**	**Old-old Adults**
Confidence: Hits			
Threat	6.68 (*.44*)	6.36 (*.53*)	5.89 (1.08)
Warning	6.59 (*.31*)	5.94 (1.21)	6.37 (*.57*)
Threat and warning	6.40 (*.73*)	6.12 (*.99*)	6.55 (*.41*)
Control	6.72 (*.21*)	6.35 (*.53*)	6.25 (*.55*)
Confidence: False Alarms to Related Lures			
Threat	5.97 (.79)	5.91 (*.96*)	5.62 (1.03)
Warning	5.69 (*.66*)	5.46 (1.26)	5.82 (1.17)
Threat and warning	5.60 (1.17)	5.48 (1.28)	6.08 (.76)
Control	5.94 (*.86*)	5.80 (*.95*)	5.42 (1.41)

Note: Standard deviations are given in parentheses.

Table 4

Average Gamma Correlations for Studied Items and Related Lures.

Measure	Young Adults	Young-old Adults	Old-old Adults
Confidence: Gamma on Studied Items			
Threat	.88 (*.14*)	.73 (*.41*)	.69 (*.26*)
Warning	.91 (*.11*)	.75 (*.33*)	.65 (*.44*)
Threat and warning	.84 (*.19*)	.73 (*.41*)	.67 (*.48*)
Control	.88 (*.14*)	.73 (*.25*)	.47 (*.69*)
Confidence: Gamma on Related Lures			
Threat	.53 (*.49*)	.63 (*.42*)	.59 (*.51*)
Warning	.46 (*.46*)	.54 (*.42*)	.42 (*.63*)
Threat and warning	.56 (*.34*)	.29 (*.64*)	.42 (*.62*)
Control	.45 (*.52*)	.61 (*.42*)	.32 (*.62*)

Note: Standard deviations are given in parentheses.

Table 5

Frequencies for categorical variables included in the multiple linear regression model.

	n	**Working**	**Retired**
Threat	43	9	34
Control	34	11	23

Table 6

Correlation table for continuous variables included in multiple linear regression.

Variable	1	2	3
1. Age			
2. Years of Education	-.092		
3. SCQ Score	.138	-.238*	

Note: *p < .05, **p < .01, ***p < .001

Table 7

Predictors included in the exploratory multiple linear regression on false alarms to related lure words.

Variable	***B***	***SE (B)***	***β***	***t***	***p***
Constant	.734	.095		7.759	.000
Retirement Status	-.131	.108	-.265	-1.215	.229
Group	.213	.119	.489	1.786	.079
Age	.000	.011	-.004	-.012	.991
Years of Education	.026	.025	.339	1.040	.302
SCQ Score	-.005	.008	-.228	-.678	.500
Group*Years of Education	.013	.019	.136	.699	.487
Group*Age	.007	.008	.181	.925	.359
Group*SCQ Score	.001	.006	.030	.143	.886
Group*Retirement Status	.301	.137	.692	2.202	.031
Age*Years of Education	.000	.002	-.043	-.289	.774
Age*Retirement Status	-.013	.012	-.362	-1.099	.276
Age*SCQ Score	-.001	.000	-.249	-1.644	.105
Retirement Status*Years of Education	-.053	.026	-.626	-2.020	.048
Retirement Status*SCQ Score	.012	.007	.472	1.650	.104
SCQ Score* Years of Education	-.001	.001	-.070	-.480	.633

Table 8

Predictors included in the final multiple linear regression on false alarms to related lure words.

Variable	***B***	***SE (B)***	***β***	***t***	***p***
Constant	.683	.072		9.554	.000
Retirement Status	-.101	.084	-.205	-1.209	.231
Group	.238	.106	.547	2.256	.027
Years of Education	-.010	.028	-.135	-.370	.713
Group*Years of Education	.098	.044	.989	2.209	.030
Group*Retirement Status	.316	.119	.727	2.650	.010
Retirement Status*Years of Education	-.010	.032	-.118	-.314	.754
Group*Years of Education*Retirement Status	-.094	.048	-.893	-1.955	.055

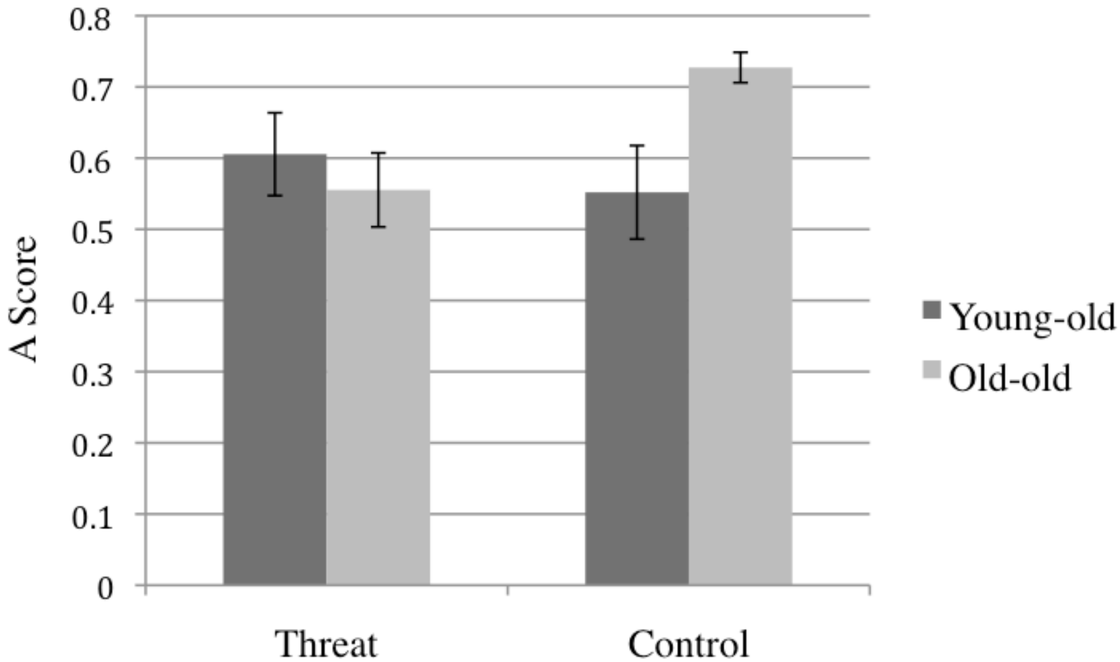
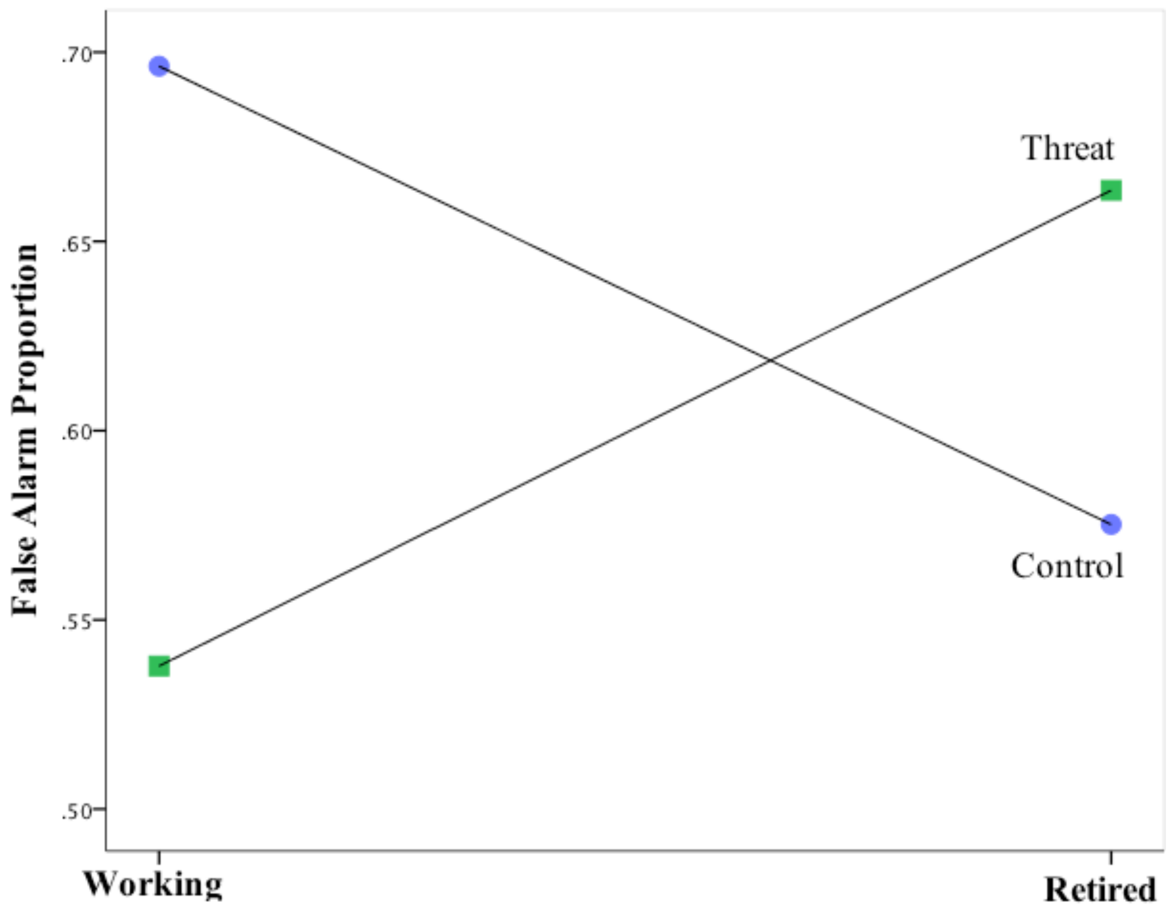

0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
A Score
Threat
Control
Young-old
Old-old

Figure 1.



.70
.65
.60
.55
.50
False Alarm Proportion
Threat
Control
Working
Retired

Figure 2.

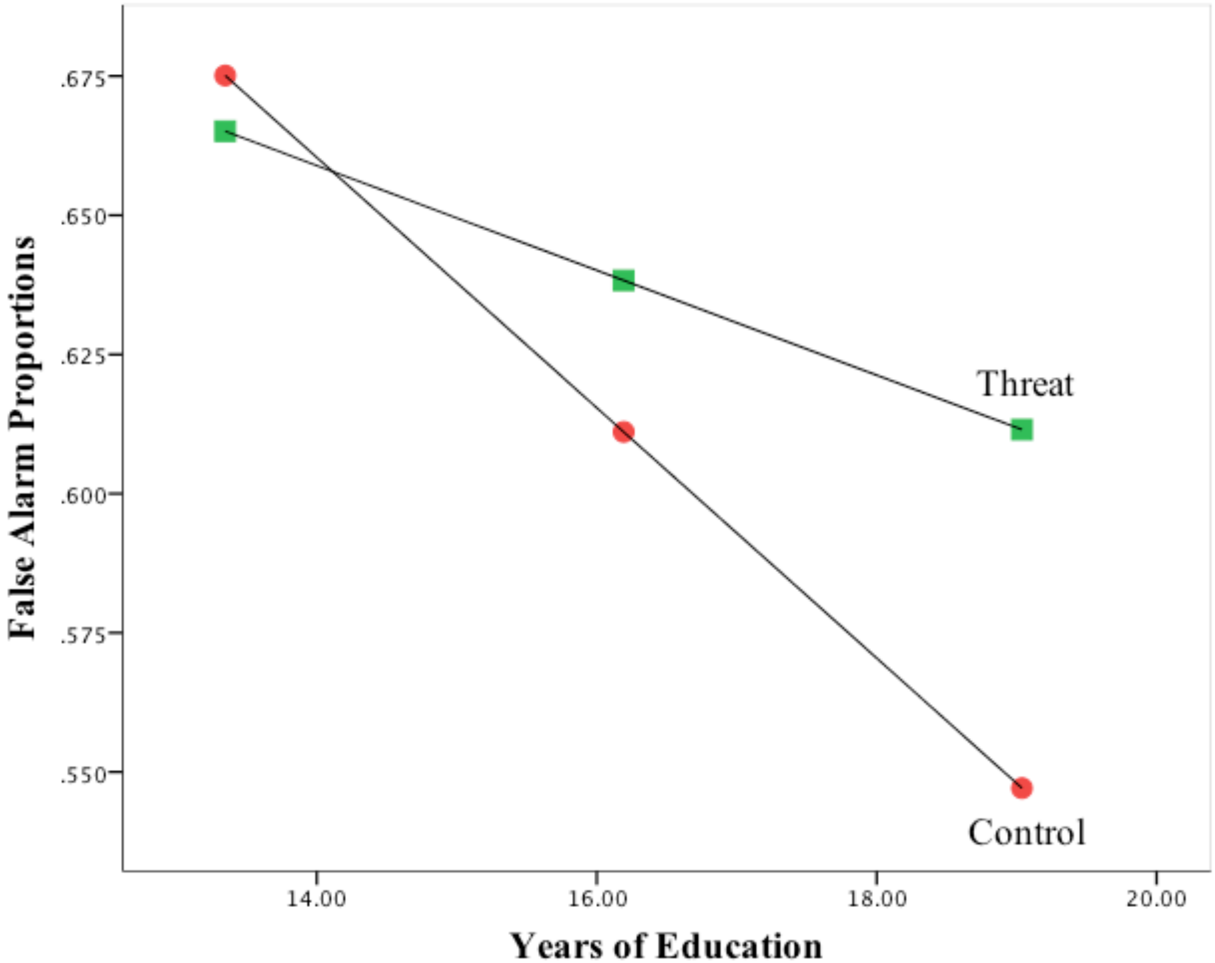

.675
.650
.625
.600
.575
.550
False Alarm Proportions
Threat
Control
14.00
16.00
18.00
20.00
Years of Education

Figure 3.