The Impact of Stereotype Threat on Age Differences in Memory Performance

Thomas M. Hess,[1] Corinne Auman,[1] Stanley J. Colcombe,[2] and Tamara A. Rahhal[3]

[1]Department of Psychology, North Carolina State University, Raleigh.
[2]Department of Psychology, University of Illinois at Urbana-Champaign.
[3]Department of Psychology, Trinity College, Hartford, Connecticut.

This study investigated the hypothesis that age differences in memory performance may be influenced by stereotype threat associated with negative cultural beliefs about the impact of aging on memory. Recall was examined in 48 young and 48 older adults under conditions varying in the degree of induced threat. Conditions that maximize threat resulted in lower performance in older adults relative to both younger adults and to older adults who did not experience threat. The degree to which threat affected older adults' performance increased along with the value that these individuals placed on their memory ability. Older adults' memory performance across experimental conditions was observed to covary with degree of activation of the negative aging stereotype, providing support for the hypothesized relationship between stereotype activation and performance. Finally, stereotype threat also influenced mnemonic strategy use, which in turn partially mediated the impact of threat on recall. These results emphasize the important role played by contextual factors in determining age differences in memory performance.

A widely held belief in our culture is that aging has a negative impact on memory functioning (cf. Hummert, 1999). Existing research provides ample empirical support for such a view, revealing that older adults perform more poorly than younger adults across a wide variety of tasks (for a review, see Zacks, Hasher, & Li, 2000). Naturally, a primary focus of research efforts has been to identify the causal mechanisms underlying such age-related variations in performance. Given the negative impact of aging on physiological structures, it is only natural that a major thrust in the study of cognitive aging has been to link memory decline to such changes (cf. Prull, Gabrieli, & Bunge, 2000); if cognition is based in brain structures, then it makes perfect sense that deterioration of such structures with aging should impact performance.

Whereas this logic is quite reasonable, the emphasis on a biological approach to the study of aging and cognition may obscure the role played by other types of factors in determining the degree to which age differences are observed in memory performance. For example, research has also demonstrated that age-related variation in task interpretation (e.g., Adams, 1991), social–emotional goals (e.g., Carstensen & Turk-Charles, 1994), and motivation (e.g., Hess, Rosenberg, & Waters, 2001) may account for age differences in both qualitative (e.g., the nature of processing) and quantitative (e.g., amount of effort expended) aspects of memory performance. Interestingly, changes in the nature of the task or in the instructions given to the individual can greatly reduce or eliminate age differences in certain contexts (Hess et al., 2001; Rahhal, Hasher, & Colcombe, 2001; Rahhal, May, & Hasher, 2002). Taken together, such findings emphasize the context specificity of age differences in memory performance while highlighting the flexibility in older adults' performance across contexts.

One potentially important contextual factor has to do with the effect of negative group stereotypes on performance. Steele (1997; Steele & Aronson, 1995) has introduced the notion of stereotype threat as one means of understanding such effects. Specifically, he argued that negative stereotypes about a group may have a detrimental impact on the behavior of group members when they are put in the position of potentially confirming that stereotype. Although the specific mechanisms mediating this effect are not entirely clear, Steele has speculated that "threat" may affect the cognitive performance of group members by raising anxiety levels and/or lowering motivation, both of which may result in low levels of performance that conform to stereotype-based expectations. Importantly, threat is hypothesized to be greatest for those who strongly identify with the stereotyped domain. In other words, group members who value the trait or ability that is being stereotyped are most likely to experience threat.

This framework was used initially to understand the impact of negative stereotypes of ability on African American students' academic performance. For example, Steele and Aronson (1995) found that African American college students performed significantly worse on an SAT-type verbal test when the study was labeled as one in which ability was to be assessed than when ability was not mentioned. They also observed evidence for implicit activation of the negative African American stereotype in those under threat. Subsequent work has both demonstrated the utility of the stereotype threat framework for understanding group differences in young adults' performance in other domains (e.g., Aronson et al., 1999; Spencer, Steele, & Quinn, 1999; Stone, Lynch, Sjomeling, & Darley, 1999) and supported the hypothesis that threat is most likely to occur in situations involving a valued trait or an important part of an individual's

identity (Leyens, Désert, Croizet, & Darcis, 2000; Stone et al., 1999).

The stereotype threat framework might also be usefully applied to understanding potential consequences of the social context on older adults' memory performance. The prevalence of negative aging stereotypes in our culture (Erber, 1999; Hummert, 1999) coupled with differential treatment of older adults in accordance with these stereotypes (e.g., Kemper, 1994; Ryan & Cole, 1990) create conditions ripe for promoting stereotype threat. For example, Hummert (1999) has argued that cues associated with older adults' physical appearance and behavior may activate aging stereotypes in others, which in turn influence their perceptions of and behavior toward these same older adults. It is likely that such stereotype-based treatment by others heightens older adults' awareness of the aging-related beliefs held by these individuals as well as the fact that they are viewed as members of the stereotyped group, both of which are necessary components for the operation of stereotype threat (Wheeler & Petty, 2001).

Work by Levy and colleagues has suggested that negative stereotypes about aging may result in cultural variations in the impact of aging on memory performance (Levy & Langer, 1994; but see Yoon, Hasher, Feinberg, Rahhal, & Winocur, 2000) as well as situational variability in the performance of older adults (Levy, 1996; Levy, Hausdorff, Hencke, & Wei, 2000). For example, differential activation of positive versus negative aging stereotypes in older adults has resulted in variations in memory performance (Levy, 1996) and physiological arousal (Levy et al., 2000). Thus, there is emerging evidence that negative stereotypes that exist about aging within our culture might not just affect others' perceptions about the abilities of older adults, but also the performance of those being stereotyped.

Whereas the just-cited work has obvious implications regarding threat effects and aging, little aging research has been conducted specifically within the stereotype threat framework. For example, the research by Levy (1996) suggests the possibility of negative and positive stereotype activation having an impact on older adults' memory performance. The fact that stereotype activation was accomplished through nonconscious means, however, raises questions as to whether the mechanisms underlying the observed effects on performance are similar to those associated with stereotype threat. Threat is thought to be associated with conscious apprehension of stereotype-relevant cues resulting in an awareness of the stereotype-related implications of the situation, including the diagnostic value of the performance context for the stereotype-relevance ability (Wheeler & Petty, 2001).

In this study, we sought specifically to examine the potential impact of stereotype threat on older adults' memory performance as well as conditions that might negate threat or enhance performance. Following the general procedure used by Aronson and colleagues (1999), stereotype threat was manipulated by informing young and older adults about research findings depicting either the negative impact of aging on memory or more optimistic results and then informing them that they would be given a memory test to further explore these findings. A control condition was also included in which no information about the impact of aging on mem-

ory was provided. Stereotype activation was then assessed, after which participants completed a free recall memory test.

In all groups, information provided about the effects of aging on memory was expected to serve as a prime for activation of positive or negative aging stereotypes. In addition, given the prevalence of negative images of aging in our culture, we expected that participants in the control condition would exhibit negative stereotype activation similar to that observed in participants receiving negative information about aging. Consistent with implications of the stereotype threat framework, it is important to demonstrate that activation of the negative stereotype occurs in those likely to experience threat. Such activation, however, is a necessary but not sufficient condition for threat to occur. That is, simply being aware of the negative stereotype should not automatically lead to threat. Stereotype threat should be most likely when individuals are put in a situation in which they are made aware of stereotype-based implications of their own group membership, they are given a task that is diagnostic of the stereotyped ability, and they personally value their ability in the stereotyped domain.

Given these conditions, we expected that the memory performance of older adults would be most susceptible to our experimental manipulations. We hypothesized that older adults in the negative information group, where the conditions for threat are greatest, would exhibit poorer memory performance than participants in the positive information group, where the conditions associated with threat have been minimized. In addition, we predicted that the impact of negative information on performance would be greatest on those older adults who valued their memory performance the most. Our expectations about the control condition were less certain. If the typical laboratory testing situation is sufficient to heighten threat, then older participants in this condition should perform like those in the negative information condition. We did try, however, to lower the operation of threat in the control condition by minimizing cues associated with expectations that might be raised under normal testing conditions (e.g., not mentioning the aging component of the task). Thus, we hypothesized that performance levels in the control group would fall somewhere between the positive and negative information conditions.

We also investigated potential mediators of threat on performance. Consistent with expectations derived from Steele (1997), we examined whether the conditions associated with threat also heightened anxiety levels, which in turn might account for performance variance associated with threat. We also examined the possibility that threat might affect performance through strategy use. When performing tasks with which the individual does not have extensive practice, individuals must formulate an appropriate strategy as well as implement it, both of which place demands on executive functions associated with working memory. To the extent that threat is associated with increases in factors such as anxiety, arousal, or intrusive/disruptive thoughts, executive functioning might suffer. Thus, we hypothesized that strategy use, as reflected in output clustering in recall, would be associated with threat and partially mediate the relation between our experimental conditions and memory performance.

The younger adults in the study were not expected to be

affected by our experimental manipulations because the previously described conditions for threat were not present for these individuals. There is some evidence in the literature for younger adults' performance being influenced by the priming of aging stereotypes (e.g., Bargh, Chen, & Burrows, 1996; Dijksterhuis, Aarts, Bargh, & van Knippenberg, 2000). Priming in these studies, however, was nonconscious. In contrast, stereotype threat is hypothesized to operate at a more conscious level as the individual recognizes the applicability and potential consequences of the stereotype in a given situation. This suggests that the performance outcomes associated with these two types of stereotype activation may arise through different mechanisms (cf. Wheeler & Petty, 2001).

METHODS

Participants

The older adult group consisted of 27 women and 21 men (M_{age} = 70.8 years, range = 62–84) recruited from the community through newspaper advertisements. They were paid $15 for their participation. The young group consisted of 26 women and 22 men (M_{age} = 19.3 years, range = 18–30) who were students in an introductory psychology course at North Carolina State University. They participated to fulfill an optional course assignment.

Tasks, Tests, and Materials

Attitudes toward memory.—Two subscales of the Metamemory in Adulthood (MIA) questionnaire (Dixon & Hultsch, 1984) were used. The Memory Achievement (MIA-Ach) subscale (α = .76 to .79), which consists of 16 items (e.g., "It is important to me to have a good memory"), was used to assess the value that individuals place on their memory ability, a hypothesized moderator of threat effects in older adults. The Memory Anxiety (MIA-Anx) subscale (α = .76 to .79) was used to assess anxiety associated with memory performance both as an outcome of threat and as a potential mediator of threat effects on memory performance. It consists of 14 items (e.g., "I get upset when I cannot remember something").

Research reports.—Four brief newspaper-type reports of research on memory and aging were constructed. Two of these discussed findings that demonstrated that older adults' memory skills were worse than those of younger adults, as in the following excerpt taken from one article:

> Although findings such as these only reinforce our most negative conceptions of aging on mental abilities, these researchers note that this does not necessarily imply that older adults are unable to function in everyday life. They suggest, however, that in order to maintain adequate levels of functioning, older adults may have to increasingly depend upon the help of memory tools as well as friends and family.

Two analogous reports discussed more positive findings regarding the relationship between aging and memory, as illustrated by the following excerpt:

> Findings such as these continue to damage our mostly negative conceptions of the effects of aging on mental abilities. Rather than supporting the view that biological changes lead to inevitable losses, these findings suggest that the degree of memory loss is to a certain extent under control of the environment and the individual.

The two positive articles were 319 and 424 words in length, whereas the negative articles were 318 and 431 words long. Each was printed in newspaper-width columns and then cut out and pasted onto a clean sheet of white paper. This paper then was reproduced using a copy machine in order to promote the perception that the articles were actual newspaper reports.

Stereotype activation.—A task similar to that of Banaji and Hardin (1996) was used to assess differential activation of positive and negative aging stereotypes in response to our threat manipulation. Participants viewed a series of 54 word pairs on a computer monitor, each consisting of a prime word followed by a trait term. In the 36 target pairs, the prime word was either young or old, with each prime being paired with 9 different positive trait terms and 9 different negative trait terms. The remaining 18 prime–trait pairs consisted of a neutral prime—blank or none—and a neutral trait word. These pairs were included as fillers to help prevent the development of response sets. Selection of trait terms was based on the Anderson's (1968) likability norms, with the 18 positive, neutral, and negative traits being selected from those ranked in the top, middle, and bottom third of likability ratings, respectively. Each of these sets of 18 traits was further divided into two sets of 9 traits each within valence type, with average syllable length and frequency of occurrence in the English language being controlled over the resulting six sets. These trait terms formed the basis for the participants' evaluative judgments during the assessment of stereotype activation. The two sets of positive and negative traits were systematically rotated through prime conditions so that the traits in each set appeared equally often with each age-related prime within each Age Group × Stereotype Condition participant group.

Participants were informed that their primary task was to categorize the second word in each pair as good or bad when it appeared on the monitor. They were told to ignore the prime word and to make their decision as quickly as possible. A voice key was used to measure the time from onset of the trait term until response. For each trial an asterisk appeared in the middle of the screen as a fixation point. After 1 s, it was replaced by a prime word, which was presented for only 200 ms. The prime was then masked, with the associated trait being displayed 100 ms later. The trait term remained on the screen until the participant responded. The next trial sequence began 1,500 ms later. The sequence of timing used for the presentation of prime–trait pairs is associated with minimization of conscious control over the impact of the prime. Prime–trait pairs were presented in random order.

The primary data obtained from this task were mean response times to positive and negative traits following the

young or old prime. Differential facilitation of judgments to positive versus negative trait terms across prime conditions (e.g., faster judgments of negative traits following the old prime than after the young prime) was assumed to reflect automatic activation of stereotype-related evaluative information associated with the two prime categories.

Memory.—The free-recall task used a list of 30 words containing five moderate to high frequency exemplars of six different semantic categories based on Howard's (1979) norms. Previous use of this task (Hess & Pullen, 1994) revealed that older adults performed as a group around the mid range (e.g., 15 words recalled) of possible performance, thereby allowing sufficient room for movement in performance as a function of our experimental manipulations. The words were typed in two columns of 15 words each on a sheet of standard typing paper. The words were randomly listed with the stipulation that two words from the same category could not appear in adjacent positions on the page.

Participants studied the word list for 2 min, after which they wrote down as many words as they could remember in the order they remembered them. A minimum of 3 min was allocated for recall. If a participant was still writing at this point, more time was allowed until 30 s had passed without an additional response. The primary measure of interest obtained from this task was the proportion of the words from the list that were correctly recalled. Strategy use was also assessed by examining the extent to which words from the same semantic category were recalled together. The adjusted ratio of clustering (ARC; Roenker, Thompson, & Brown, 1971) was used for this purpose. This measure has the desirable characteristics of controlling for chance clustering levels and being relatively independent of level of recall.

Background measures.—Vocabulary Test II from the Kit of Factor-Referenced Tests (Ekstrom, French, Harman, & Derman, 1976) was used to assess verbal ability. The WAIS III Letter–Number Sequencing task (Wechsler, 1997) was used to assess working-memory functioning, and Salthouse and Coon's (1994) letter and pattern comparison tasks were used to assess processing speed. The SF-36 Health Survey (Ware, 1993) was used to obtain summary scores relating to both self-rated physical and mental health (αs = .88 to .93). Scores from these tests were collected for use as possible covariates to the extent that these potentially influential participant characteristics varied across experimental conditions.

Procedure

Sixteen participants within each age group were randomly assigned to one of three different stereotyping conditions: positive, control, or negative. Within the week prior to testing, the older adults were sent the MIA-Ach scale, which they were instructed to complete and bring to the test session. Due to scheduling complexities, younger adults completed the MIA-Ach upon their arrival at the lab (see Appendix, Note 1). At the beginning of the session, participants in the positive and negative stereotyping conditions were informed about recent scientific evidence that either contradicted (positive condition) or reinforced (negative condition) traditional views regarding the effects of aging on memory.

They were told that the present study was concerned with further understanding this evidence. They then read one of the "newspaper" articles appropriate to their condition, which, they were informed, related to the just-mentioned research findings. Following this, participants completed an unrelated questionnaire as a filler task and then read the second article in order to further reinforce the aging-related perspective highlighted in the first. Those in the control condition were not presented with any of this information. Instead, they were informed that they would be completing a series of tasks, the first of which was a decision-making task.

All participants were given the stereotype activation assessment task next, followed by the free recall task, the MIA-Anx scale, the SF-36 Health Survey, and the three cognitive ability tests. Debriefing followed, during which the nature of the newspaper articles was explained and the findings reported in them were put into the larger context of research on aging and memory.

RESULTS

Preliminary Analyses

A series of 2 × 3 (Age Group × Stereotype Condition) analyses of variance (ANOVA) were conducted on the background measures to test for inadvertent confounds associated with participant assignment to experimental conditions. As seen in Table 1, the associated age effects are consistent with those typically observed in studies of aging and cognition. The only condition-related effect was a significant Age × Condition interaction for prescription drugs, $F(2,90) = 4.55, p < .02$, due to medication rates in the older group being greater in the control condition than in the other two conditions. Inclusion of this measure as a covariate in the primary analyses, however, did not alter the results. The absence of any other effects indicates that our random assignment was successful in controlling for factors that could potentially confound observed condition effects. Because the age range of our older group was quite large (22 years) and age is associated with memory, the mean age of participants was also compared across conditions within this group. There was no systematic variation in age ($F < 1$), with the mean ages in the positive, control, and negative conditions being 70.8, 68.9, and 71.4, respectively.

Table 1. Age Differences in Sample Characteristics

Measure	Young		Old			
	M	SD	M	SD	F(1,90)	p
Years of education	12.9	1.1	15.6	2.9	38.65	<.001
No. of prescription drugs	0.9	1.0	3.3	2.8	38.39	<.001
SF-36: Physical health	53.0	4.9	44.3	8.9	34.85	<.001
SF-36: Mental health	45.6	10.4	54.2	8.9	19.11	<.001
Speed	0.71	0.65	−0.71	0.52	137.25	<.001
Letter–number sequencing	13.2	3.2	9.9	2.2	32.02	<.001
Vocabulary	27.1	2.5	31.3	2.6	66.76	<.001

Notes: Score ranges: Letter–number sequencing = 0–21; Vocabulary = 0–36. Speed is the mean *z* score for the letter and pattern comparison tasks. SF-36 scores are *T* scores.

Main Analyses

Our primary interest was in the impact of the stereotype threat manipulation on (a) stereotype activation, (b) memory performance, (c) strategic behavior during recall, and (d) anxiety about memory. The influence of age group, stereotype condition, and the value placed on memory functioning on the dependent variables associated with each of these factors was examined using GLM-based ANOVAs, with all variables between participants. Age group and condition were treated as categorical variables using effect coding, whereas the MIA-Ach subscale score was entered into all analyses as a continuous variable centered at 0 through standardization.

Stereotype activation.—Response times from the trait judgment task were used to assess stereotype activation. Of primary interest was the speed with which participants made judgments about positive and negative traits following the "old" prime. Individual response times were eliminated due to voice key malfunction or if they were 3 standard deviations above or below a participant's mean for all responses. This resulted in the exclusion of 3.2% of the target response times in the young group and 4.8% in the older group. Mean response times were then calculated from the remaining times and examined using an Age Group × Stereotype Condition × MIA-Ach × Prime × Trait Valence ANOVA, with the last two variables within participants. (MIA-Ach was included to be consistent with subsequent analyses; we did not expect it to affect performance.) Here and in the remaining analyses we discuss only those effects relevant to our research questions.

Of primary interest was the anticipated Condition × Prime × Valence interaction, which approached significance, $F(2,84) = 3.01, \eta^2 = .07, p = .054$. Importantly, this effect was not moderated by age ($F < 1$), suggesting that the obtained stereotype activation effects were not specific to either age group. To examine specific predictions regarding activation, separate ANOVAs were conducted within each prime condition. The pattern of responses to positive and negative traits following the young prime did not vary across conditions ($F < 1$, Table 2). This finding was anticipated given that there was no reason to expect that perceptions about young adults would be affected by our stereo-

type manipulation that was directed at altering activation of aging stereotypes. In contrast, the hypothesized Condition × Valence interaction was significant in the old prime condition, $F(2,84) = 5.18, p = .01$, with responses to positive and negative traits varying in the hypothesized manner across conditions (Table 2). Focused between-condition contrasts of the valence effect indicated that the pattern of response times did not vary across the control and negative conditions ($F < 1$). In contrast, the pattern of response times in the positive condition differed from that in both the control, $F(1,84) = 5.70, p = .02$, and negative, $F(1,84) = 9.40, p = .003$, conditions. In both cases, relative to the positive condition, responses to positive traits were slower and those to negative traits were faster. This pattern of responding suggests that we were successful in manipulating the activation of positive versus negative aging stereotypes. The similarity between responses in the control and negative conditions also suggests that, in the absence of active attempts to manipulate stereotype activation, the negative aging stereotype is dominant.

Recall.—The proportion of words from the list that were correctly recalled was examined next. Our interest here was in the differential impact of context on performance across age groups, and thus we wanted to make sure that any observed variation was not due to age differences in cognitive ability factors that might independently influence memory performance. To do so, we first obtained a general measure of cognitive ability by averaging standardized scores for the letter–number sequencing tasks and the speed tasks. Both tasks tap into basic cognitive mechanisms thought to affect more complex functions, and performance on the two tasks was highly correlated, $r(94) = .60, p < .001$. Preliminary analyses also revealed that this ability measure was significantly related to both age and recall performance, but did not interact with any other variable in predicting recall, thus satisfying conditions for its use as a covariate in an analysis of covariance (ANCOVA).

Initial examination of the results from an Age × Condition × MIA-Ach ANCOVA on recall revealed an outlier in the older control group that exceeded accepted leverage levels (Stevens, 1996). Analyses on the remaining data indicated that younger adults (adjusted $M = .60, SD = .11$) recalled more than the older adults (adjusted $M = .53, SD = .17$), $F(1,82) = 4.12, \eta^2 = .05, p = .05$. In addition, recall varied reliably across conditions, $F(2,82) = 3.69, \eta^2 = .03, p = .03$, and was negatively associated ($r = -.26$) with memory achievement scores, $F(1,82) = 5.32, \eta^2 = .02, p = .02$. A significant Age × MIA-Ach interaction was also obtained, $F(1,82) = 7.44, \eta^2 = .08, p = .01$, due to the impact of achievement scores being reliable for the older adults ($r = -.31, p = .04$), but not for the younger adults ($r = .08, p = .58$). The predicted three-way interaction only approached significance, $F(2,82) = 2.47, \eta^2 = .06, p = .09$, most likely due to the low observed power (.48) associated with this effect in the present analysis. Planned analyses within age groups, however, clearly indicated that the condition effects were specific to the older adults.

As expected, examination of the young adults' data revealed no significant effects. In contrast, a significant main

Table 2. Mean Response Times (in Milliseconds) for Positive and Negative Traits Across Prime and Stereotype Conditions

Condition	Young Prime		Old Prime	
	Positive Traits	Negative Traits	Positive Traits	Negative Traits
Positive				
M	788	932	759	908
SD	163	209	151	204
Control				
M	777	901	762	867
SD	152	179	142	153
Negative				
M	790	950	809	902
SD	191	237	196	193

effect was obtained for condition in the older group, $F(2,40) = 4.12$, $\eta^2 = .17$, $p = .02$, with Tukey's HSD test revealing that the locus of the effect was due to the significantly ($ps < .05$) lower level of recall in the negative condition when compared with the other two (Table 3), which did not differ from each other ($p = .7$). This effect must be interpreted with caution, however, given the significance of the critical Condition $\times$ MIA-Ach interaction, $F(2,40) = 3.33$, $\eta^2 = .14$, $p = .05$. Consistent with expectations, this effect reflects the fact that the impact of our threat manipulation increased as the value placed on memory functioning increased. Specifically, there was little relation between recall and MIA-Ach scores in the positive condition ($r = -.10$, $p = .36$), but reliable negative relations were observed in the control ($r = -.50$, $p = .03$) and negative ($r = -.70$, $p = .001$) conditions. This finding is consistent with the notion that increases in stereotype threat will have the greatest impact on those who value the threatened skill most (see Appendix, Note 2).

The impact of our stereotype threat manipulation on the determination of age differences can also be seen when age effects are examined within each condition. No statistically reliable differences were observed between the young and older adults in the positive condition ($ps > .12$). In contrast, older adults had significantly poorer recall than younger adults in the negative condition, $F(1,27) = 5.29$, $\eta^2 = .16$, $p = .03$, and the Age Group $\times$ MIA-Ach interaction was significant in both the control condition, $F(1,26) = 9.52$, $\eta^2 = .27$, $p = .01$, and the negative condition, $F(1,27) = 4.33$, $\eta^2 = .14$, $p = .05$. These latter two effects reflect the previously described negative relationships between MIA-Ach scores and recall and the differential predictive power of these scores across age groups.

Clustering.—We next investigated strategy use during recall. Initial analyses once more revealed that conditions for use of our cognitive ability measure as a covariate were satisfied, and thus ARC scores were examined using an Age Group $\times$ Condition $\times$ MIA-Ach ANCOVA. The only significant effect obtained from this analysis was a main effect of condition, $F(2,83) = 4.70$, $\eta^2 = .10$, $p = .01$, due to ARC scores being significantly ($ps < .04$) greater in the positive (adjusted $M = .61$, $SD = .33$) and control (adjusted $M = .64$, $SD = .28$) conditions than in the negative condition (adjusted $M = .43$, $SD = .30$).

Although the three-way interaction was not significant ($p =$

.25), the observed power for this effect was rather low (.30) and planned analyses were once again conducted within age groups to test specific hypothesis. For the young group, the only significant effect was due to condition, $F(2,41) = 5.70$, $\eta^2 = .22$, $p = .01$, reflecting the unexpected finding that clustering in the negative condition was significantly ($ps < .05$) lower than that in either the positive or control conditions (Table 3). It is unclear why this effect was obtained, particularly in the absence of variations in recall across conditions in this age group, suggesting the effect is spurious in nature. For the older adults, we once again obtained a significant interaction between condition and achievement, $F(2,41) = 3.38$, $\eta^2 = .14$, $p = .04$. This effect was due to the differential relationship between ARC and MIA-Ach scores across conditions. A significant negative relationship was observed in the negative condition ($r = -.57$, $p = .01$), whereas a marginally significant positive relationship was observed in the positive condition ($r = .43$, $p = .10$). A negative, but nonsignificant relationship was obtained in the control condition ($r = -.16$, $p = .55$). Comparisons across conditions revealed that the correlations in the positive and negative condition were significantly different ($Z = 2.82$, $p = .005$), whereas neither of these correlations differed from that in the control condition ($ps > .11$).

Interestingly, recall performance by the older adults in all three conditions was positively associated with the degree of clustering (positive $r = .14$; control $r = .68$; negative $r = .81$), although the correlations were only statistically reliable in the control and negative conditions ($ps < .001$). This finding along with the observed variations across conditions in the relationship between clustering and memory achievement is consistent with our hypothesis that stereotype threat affects memory performance through its impact on strategy use. This mediational relationship can be demonstrated statistically by controlling for clustering when examining recall. The two effects reflective of the hypothesized threat effects on memory condition and Condition $\times$ MIA-Ach accounted for 29% of the variance in recall, $F(4,40) = 4.97$, $p < .01$, after controlling for cognitive ability and MIA-Ach. Subsequent control for ARC scores reduced the recall variance accounted for by these two effects to 12%, $F(4,39) = 3.12$, $p < .05$, and the Condition $\times$ MIA-Ach interaction was no longer significant ($p = .47$). Thus, approximately 58% of the variance associated with stereotype threat-related effects in recall was accounted for by strategy use.

Anxiety.—Our final analysis examined the impact of stereotype threat on anxiety about memory using the MIA-Anx scores as dependent measures. No significant effects were observed with the variable. In addition, no significant correlations were observed between MIA-Anx scores and recall and clustering scores. This lack of significant effects precluded the examination of anxiety as a potential mediator of threat.

DISCUSSION

This study examined the impact of contextual factors in determining older adults' memory performance and, concomitantly, the magnitude of age differences therein. Specifically, consistent with Steele's (1997) notion of stereotype threat,

Table 3. Adjusted Mean Recall and Clustering as a Function of Age and Condition

Age Group	Recall			ARC Scores		
	Positive	Control	Negative	Positive	Control	Negative
Young						
M	.61	.61	.59	.58	.60	.25
SD	.11	.12	.12	.33	.28	.26
Old						
M	.58	.56	.44	.64	.68	.61
SD	.14	.17	.17	.25	.29	.25

we hypothesized that the memory performance of older adults would be negatively affected when environmental cues activate negative expectations about the memory abilities of older adults. Conversely, when such cues are eliminated, it was expected that older adults' performance would be optimized. The results of our research were generally consistent with expectations, providing potentially important insights concerning the factors underlying observed age differences in memory performance.

To demonstrate the impact of stereotype threat on older adults' memory performance, evidence for differential stereotype activation in response to relevant situational cues is necessary. Such evidence was obtained in our study. Specifically, when participants were told the study was examining aging-related memory decrement, their judgments about negative trait terms were faster and those about positive traits were slower following brief exposure to the word old than when they were told that the study was examining positive aspects of aging. This differential facilitation of responses for positive versus negative traits following the old prime across conditions is suggestive of differential activation of positive and negative aging stereotypes in response to exposure to stereotype relevant cues. Interestingly, the same pattern of responses was observed when the control condition was contrasted with the positive condition, suggesting that the evaluative content associated with the term old is relatively negative in both young and older adults without any explicit prompting.

Although activation of negative stereotype information is a necessary condition for stereotype threat to occur, it is not thought to be sufficient to negatively influence performance. Additional conditions, such as self-relevance of the negative stereotype (i.e., one must be a member of the group), the assessment context being perceived as diagnostic of the stereotyped ability, and high importance attached to this ability, should also be present for performance to be affected. This led to the expectation that variations in the salience of information about the impact of aging on memory would only affect performance in the older adult group, with those who place high value on their memory ability being affected most. This was, in fact, the pattern of performance observed.

Younger adults, for whom aging stereotypes were not applicable, did not exhibit significant variation in recall across experimental conditions. In contrast, older adults' performance varied as a function of condition and the value placed on memory in a manner suggestive of the operation of stereotype threat in the negative condition and, to a lesser degree, in the control condition. Specifically, consistent with expectations derived from the stereotype threat framework (Steele, 1997), performance was most negatively affected when older adults were exposed to threat and when they also had some personal investment in their memory skills. Thus, recall in the negative information condition was negatively correlated with the degree to which older adults reported that achievement in the memory domain was important to them. This same relation was not significant in the positive condition. Importantly, memory achievement and recall were also negatively correlated in the control condition, suggesting that threat may be operative even when explicit linkages with the aging stereotype are not made. Al-

though its strength was not as great as that in the negative condition, the presence of this relationship in the control condition points to the possibility of threat being operative under normal testing circumstances, especially in older adults who value their memory ability. This can be seen more clearly by comparing predicted recall scores for those in the positive and control conditions obtained from functions in which recall is regressed on MIA-Ach scores. Predicted scores for those with mean levels of memory achievement were .55 in the positive condition and .54 in the control condition. In contrast, predicted performance for those with relatively high memory achievement scores (i.e., 1 standard deviation above the sample mean) was .53 in the positive condition versus .44 in the control condition.

We were also interested in understanding the processes underlying the impact of our threat-related manipulations on performance. We first examined whether threat might be related to anxiety about memory, but scores on the MIA Memory Anxiety subscale were unrelated to the threat manipulation. The lack of an observed empirical relationship, however, may have to do with the MIA anxiety measure being more reflective of trait anxiety than state anxiety. The latter might be more susceptible to situational variations in stereotype threat, as suggested by recent studies by Auman (2002) and Andreoletti and Lachman (2002).

Strategy use, as reflected in clustering by semantic category during recall, was also examined as a potential mediator. Clustering is associated with successful recall of categorically organized lists, and disruption of this type of strategy use might help explain the relationship between threat and performance. Consistent with this mediational hypothesis, ARC scores were observed to vary in a manner similar to that observed for recall. Specifically, clustering declined for older adults in the negative condition as investment in memory increased. In contrast, a marginally significant increase was observed in clustering with achievement in the positive condition. This finding is once more consistent with the hypothesis that threat will have the greatest negative impact on those who value the threatened ability most. A negative relationship between achievement and clustering scores was also observed in the control condition, but the correlation was not significant. The facts that this relationship was negative and also fell between that observed in the positive and negative conditions again provide evidence that the normal laboratory testing situation might induce stereotype threat.

Subsequent analyses also demonstrated that clustering scores accounted for over half of the variance in older adults' performance attributable to threat-related effects. In other words, the impact of threat on memory performance appears to be partially mediated by strategy use, a conclusion consistent with recent findings by Quinn and Spencer (2001). A question of interest then concerns the nature of the relationship between threat and strategy use. Steele and Aronson (1995) had discussed both anxiety and motivation as being possible mechanisms through which stereotype threat affects performance, and both could conceivably be used to explain the observed relationships. For example, threat may disrupt memory performance by increasing anxiety/arousal or disruptive thoughts, which in turn may negatively

affect executive functions controlling strategy use. Alternatively, stereotype threat might affect strategy use by lowering expectations, which in turn has been shown to reduce effort and motivation (e.g., Bandura, 1977). It seems quite possible that both types of processes might be operative, or that the potential mediators might vary depending upon the situation. As noted before, the use of situation-specific measures of these factors (e.g., state anxiety) would facilitate our understanding of the mechanisms underlying this relationship.

The results of this research are noteworthy in several ways. First, support was provided for the hypothesis that stereotypes influence the behavior of older adults. The present research extends previous findings that have shown a relationship between stereotype activation and performance (e.g., Levy, 1996; Levy et al., 2000) by demonstrating a link between explicit environmental cues and both stereotype activation and performance. The observation of these effects in response to conscious experiences of the individual provides more direct support for the potential operation of stereotype threat in everyday life.

Put in a more positive light, the research also demonstrated the facilitative impact on older adults' memory when threat-related factors are minimized by giving participants information that suggested that memory decline was not inevitable. This result fits into a larger body of research that has shown that restructuring of beliefs about control over ability (e.g., Aronson, Fried, & Good, 2001; Lachman, Weaver, Bandura, Elliott, & Lewkowicz, 1992) can facilitate performance. Although such findings suggest that control-related beliefs may be important in the operation of stereotype threat-based effects, it does not appear that the manipulation of such beliefs is the only way to induce or eliminate threat. For example, varying entity-based beliefs has been shown to have both positive and negative effects on performance in the same domain (e.g., Shih, Pittinsky, & Ambady, 1999; Stone et al., 1999). It does seem obvious, however, that the manipulation of control beliefs may be an effective way to eliminate stereotype threat.

A second significant aspect of the present research is the identification of a contextual influence on older adults' memory performance. This result adds to a growing list of empirical findings (e.g., Adams, 1991; Carstensen & Turk-Charles, 1994; Hess et al., 2001; Rahhal et al., 2001) that demonstrate the impact of specific situational factors in determining adult age differences in memory performance. Such findings emphasize the importance of adopting a multidimensional approach to the study of aging and memory that considers factors beyond presumed aging-related changes in physiological structures.

Finally, our observation of the operation of stereotype threat in the memory performance of older adults has implications for both everyday functioning and our laboratory-based assessments of ability. With respect to everyday functioning, older adults in our culture are frequently exposed to reminders about the negative aspects of aging through the media, advertisements, and the behaviors of others toward them. Such environmental cues are likely to activate negative aging stereotypes in the older perceiver, potentially resulting in performance decrements or disengagement from the task domain due to threat. The operation of stereotype threat also has important implications for our assessments of age differences in memory performance to the extent that circumstances surrounding laboratory-based testing of older adults cue negative aging stereotypes. For example, ads used to recruit research participants for research on aging and the instructions given in studies of memory (Rahhal et al., 2001) may activate negative stereotypes, resulting in older adults feeling threatened in such testing situations. The differential impact of such situation cues across age groups has the potential to skew research findings and lead to inaccurate conclusions about older adults' abilities and associated age differences in performance. The negative condition in this study might be viewed as an extreme example of how laboratory instructions might depress older adults' performance, but the performance of older adults in our control group suggests that even subtle environmental cues might induce threat. Such findings emphasize both the important impact that factors associated with the assessment context can have on performance and the necessity for researchers to consider the role that such factors play in their own studies of aging and memory.

REFERENCES

Adams, C. (1991). Qualitative age differences in memory for text: A life-span developmental perspective. *Psychology and Aging, 6,* 323–336.

Anderson, N. H. (1968). Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology, 9,* 272–279.

Andreoletti, C., & Lachman, M. (2002). *Effects of stereotype information on memory performance in adulthood: Education makes a difference.* Manuscript submitted for publication.

Aronson, J., Fried, C. B., & Good, C. (2001). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 37,* 1–13.

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35,* 29–46.

Auman, C. (2002, April). *Stereotype threat influence in a medical setting.* Poster session presented at the Cognitive Aging Conference, Atlanta, GA.

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science, 7,* 136–141.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavior change. *Psychological Review, 84,* 191–215.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71,* 230–244.

Carstensen, L. L., & Turk-Charles, S. (1994). The salience of emotion across the adult life span. *Psychology and Aging, 9,* 259–264.

Dijksterhuis, A., Aarts, H., Bargh, J. A., & van Knippenberg, A. (2000). On the relation between associative strength and automatic behavior. *Journal of Experimental Social Psychology, 36,* 531–544.

Dixon, R. A., & Hultsch, D. F. (1984). The metamemory in adulthood (MIA) instrument. *Psychological Documents, 14,* 3.

Ekstrom, R. B., French, J. W., Harman, H., & Derman, D. (1976). *Kit of factor-referenced cognitive tests* (rev. ed.). Princeton, NJ: Educational Testing Service.

Erber, J. (1999). Age and memory: Perceptions of forgetful young and

older adults. In T. M. Hess & F. Blanchard-Fields (Eds.), *Social cognition and aging* (pp. 197–217). San Diego, CA: Academic Press.

Hess, T. M., & Pullen, S. M. (1994). Adult age differences in informational biases during impression formation. *Psychology and Aging, 9*, 237–250.

Hess, T. M., Rosenberg, D. C., & Waters, S. J. (2001). Motivation and representational processes in adulthood: The effects of social accountability and information relevance. *Psychology and Aging, 16*, 629–642.

Howard, D. V. (1979). *Category norms for adults between the ages of 20 and 80*. Technical Report No. NIA-79–1, Department of Psychology, Georgetown University, Washington, DC.

Hummert, M. L. (1999). A social cognitive perspective on age stereotypes. In T. M. Hess & F. Blanchard-Fields (Eds.), *Social cognition and aging* (pp. 175–196). San Diego, CA: Academic Press.

Kemper, S. (1994). "Elderspeak": Speech accommodations to older adults. *Aging and Cognition, 1*, 17–28.

Lachman, M. E., Weaver, S. L., Bandura, M., Elliott, E., & Lewkowicz, C. J. (1992). Improving memory and control beliefs through cognitive restructuring and self-generated strategies. *Journal of Gerontology: Psychological Sciences, 47*, P293–P299.

Levy, B. (1996). Improving memory in old age by implicit self-stereotyping. *Journal of Personality and Social Psychology, 71*, 1092–1107.

Levy, B., & Langer, E. (1994). Aging free from negative stereotypes: Succesful memory in China and among the American deaf. *Journal of Personality and Social Psychology, 66*, 989–997.

Levy, B. R., Hausdorff, J. M., Hencke, R., & Wei, J. Y. (2000). Reducing cardiovascular stress with positive self-stereotypes of aging. *Journal of Gerontology: Psychological Sciences, 55B*, P205–P213.

Leyens, J., Désert, M., Croizet, J., & Darcis, C. (2000). Stereotype threat: Are lower status and history of stigmatization preconditions of stereotype threat? *Personality and Social Psychology Bulletin, 26*, 1189–1199.

Prull, M. W., Gabrieli, J. D. E., & Bunge, S. A. (2000). Age-related changes in memory: A cognitive neuroscience perspective. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 91–153). Mahwah, NJ: Erlbaum.

Quinn, D. M., & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues, 57*, 55–71.

Rahhal, T. A., Hasher, L., & Colcombe, S. J. (2001). Instructional manipulations and age differences in memory: Now you see them, now you don't. *Psychology and Aging, 16*, 697–706.

Rahhal, T. A., May, C. P., & Hasher, L. (2002). Truth and character: Sources that older adults can remember. *Psychological Science, 13*, 101–105.

Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin, 76*, 45–48.

Ryan, E. B., & Cole, R. L. (1990). Evaluative perceptions of interpersonal communication with elders. In H. Giles, N. Coupland, & J. Wiemann (Eds.), *Communication, health, and the elderly*: *Fulbright Series No. 8* (pp. 172–190). Manchester, England: Manchester University Press.

Salthouse, T. A., & Coon, V. E. (1994). Interpretation of differential deficits: The case of aging and mental arithmetic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1172–1182.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 80–83.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4–28.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*, 613–629.

Steele, C. M., & Aronson, J. (1995). Contending with a stereotype: African-American intellectual test performance and stereotype threat. *Journal of Personality and Social Psychology, 69*, 797–811.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on Black and White athletic performance. *Journal of Personality and Social Psychology, 77*, 1213–1227.

Ware, J. E., Jr. (1993). *SF-36 Health Survey.* Boston: The Health Institute, New England Medical Center.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—Third Edition*. New York: Psychological Corporation.

Wheeler, S. C., & Petty, R. E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin, 127*, 797–826.

Yoon, C., Hasher, L., Feinberg, F., Rahhal, T. A., & Winocur, G. (2000). Cross-cultural differences in memory: The role of culture-based stereotypes about aging. *Psychology and Aging, 15*, 694–704.

Zacks, R. T., Hasher, L., & Li, K. Z. H. (2000). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 293–357). Mahwah, NJ: Erlbaum.

Appendix

Notes

1. The difference in times at which young and older participants completed the MIA-Ach may be a potential confound in interpretation of the results. It can reasonably be argued, however, that assessing the degree to which one values one's memory ability just prior to memory testing would prime memory concerns and amplify any threat-related effects, potentially providing a strong test of the age-group specificity of the hypothesized stereotype effects.

2. The range of education in the older group (7–21 years) was much greater than that in the younger group (12–16 years), leading to the possibility that the aging-related effects were due mainly to performance of those with either high or low levels of education. Inclusion of years of formal education in this analysis, however, did not alter the outcomes nor result in any significant effects involving this variable.