EFFECTS OF AGE-BASED STEREOTYPE THREAT AND DEMENTIA-RELATED DIAGNOSIS THREAT ON OLDER ADULTS’ OBJECTIVE AND SUBJECTIVE COGNITIVE FUNCTIONING

by

JOIE MOLDEN

B.M., University of Montevallo, 2008

B.S., University of Montevallo, 2011

M.A., University of Colorado Colorado Springs, 2014

A dissertation submitted to the Graduate Faculty of the University of Colorado Colorado Springs in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Psychology

2018
This dissertation for the Doctor of Philosophy degree by

Joie Molden

has been approved for the

Department of Psychology

by

Molly Maxfield, Chair

Brandon Gavett

Lori James

Thomas Pyszczynski

Michael Greher

Date 08/31/2017
Molden, Joie (Ph.D., Psychology)

Effects of Age-Based Stereotype Threat and Dementia-Related Diagnosis Threat on Older Adults’ Objective and Subjective Cognitive Functioning

Dissertation directed by Associate Professor Molly Maxfield

ABSTRACT

Alzheimer’s disease and other dementias represent a significant health concern. Neuropsychologists commonly see older adult patients referred to assess whether perceived cognitive changes represent normal age-related changes or decline attributable to disease processes. Effective assessment requires recognition of the part a patient’s sociocultural context plays in their functioning. For older adults, this entails understanding “stereotype threat,” whereby performance on cognitive testing may be negatively impacted by being at risk of confirming negative stereotypes of older adults as cognitively impaired. Related to stereotype threat, “diagnosis threat” (impaired performance due to expectations regarding a diagnosis) has been proposed as a possible mechanism for poorer test performance among patients with a history of mild traumatic brain injury and may also impact an individual’s perception of cognitive changes. No published studies have explored diagnosis threat among older adults, although research on dementia worry suggests that for some, concern about developing dementia negatively impacts well-being (and could include high perceived risk of dementia). A sample of 103 older adults ($M_{age} = 72.38$, $SD = 6.68$) was randomly assigned to (a) an age-based stereotype threat manipulation, (b) a dementia-related diagnosis threat manipulation, or (c) a neutral control condition. Although it was hypothesized that those in the stereotype and diagnosis threat groups would perform more poorly on cognitive testing,
performance on standard neuropsychological tests was not significantly affected by the experimental manipulation. While the manipulation also did not affect most subjective variables (including self-reported change in cognitive functioning, dementia worry, and self-rated effort), perceived stereotype threat was significantly higher among those in the age-based stereotype threat group. Correlational data suggested significant relationships between subjective variables of interest, including self-reported cognitive change, dementia worry, perceived stereotype threat, perceived risk of developing dementia, and participant perception of test performance. Implications for future research are considered.

*Keywords*: stereotype threat, diagnosis threat, aging, older adults
ACKNOWLEDGEMENTS

It seems impossible to have completed an undertaking as monumental as a dissertation—and the fact of the matter is, reaching this milestone would have been impossible were it not for the contributions of several individuals whose consistent willingness to help me (in this and other endeavors) continues to humble and move me. To Molly Maxfield, I thank you for five years of mentorship, guidance, and support not only in research endeavors, but also in professional development and personal growth—your kindness, humor, and reliability are only some of the qualities that make you an excellent mentor that I was fortunate to have. My sincerest gratitude also to Brandon Gavett, Lori James, Tom Pyszczynski, and Michael Greher for their willingness to invest their time, energy, and enthusiasm into this project as committee members; the final product was made infinitely better by their ideas, discussions, and suggestions. And of course, there is no measure of the debt I owe to my family. To my parents, Bambi and Richard, for over 30 years of ebullient and unwavering support of their perpetual-student daughter (they came to every musical performance and probably would have shown up to my dissertation defense if I’d shared the address with them); and to my partner, Stephen, for (among so many other things) the patience, generosity, and laughter he has continuously provided through each and every grad school hurdle—thank you.
# TABLE OF CONTENTS

CHAPTER

I. INTRODUCTION ........................................................................................................1
   Stereotype Threat .................................................................................................3
   Stereotype Threat and Older Adults .................................................................6
   Diagnosis Threat ..............................................................................................21
   Dementia Worry ...............................................................................................28
   The Present Study ...........................................................................................31

II. METHOD ..............................................................................................................38
   Participant Characteristics .............................................................................38
   Materials ...........................................................................................................40
   Self-report Measures ......................................................................................51
   Research Design and Procedure .................................................................57

III. RESULTS ..........................................................................................................60
   Planned Statistical Analyses .........................................................................60
   Participant Differences Between Groups ...................................................62
   Internal Consistency .......................................................................................62
   Manipulation Check .......................................................................................62
   Analyses – Hypotheses 1-4 ...........................................................................64
   Analysis – Hypothesis 5 ..................................................................................68
   Analysis – Hypothesis 6 ..................................................................................68
LIST OF TABLES

TABLE

1. Demographic Characteristics by Condition ...................................................41
2. Accuracy of Multiple Choice Manipulation Check by Condition ..................63
3. Accuracy of Debriefing Manipulation Check by Condition .........................64
5. Means, Standard Deviations, and Ranges of Subjective Variables ...............70
6. Logistic Regression Results Using All Participants (N = 101) .......................76
7. Logistic Regression Results Excluding Outliers (n = 96) ............................77
8. Correlations Between Cognitive and Demographic Variables ....................87
9. Correlations Between Subjective Variables, Cognitive Composite Measure, and Demographic Variables .................................................................88
CHAPTER I

INTRODUCTION

Demographic shifts in both the United States and other developed countries across the world have resulted in older adults composing an increasingly substantial proportion of the population (Aldwin & Gilmer, 2013). Along with this increase in representation comes an increase in chronic diseases associated with aging, including diabetes, cardiovascular disease, and cancer. But among chronic illnesses, the threat that may represent the most salient concern for aging individuals is that of dementia. Over 5 million Americans currently live with Alzheimer’s disease (AD), and estimates place it somewhere between the third and sixth leading cause of death in the United States (Alzheimer’s Association, 2016; James et al., 2014). Other dementia processes, including vascular dementia, dementia with Lewy bodies, Parkinson’s disease dementia, and other rarer dementia subtypes also represent significant health concerns for aging populations (Alzheimer’s Association, 2016; Lewy Body Dementia Association, 2014). Considering the risk for a loss of functional independence and cognitive functioning inherent in AD and other dementias, anxiety regarding such a diagnosis is a natural reaction. Indeed, AD consistently ranks as one of the most feared diseases among adults in the United States and internationally (Harvard School of Public Health & Alzheimer Europe, 2011; MetLife, 2011).

The field of neuropsychology plays an important role in helping older adults and their families differentiate between whether an individual’s cognitive symptoms represent
dementia or simply the effects of normal aging—which include declines in abilities such as processing speed, memory, executive functioning, and some aspects of attention, visuospatial, and language functioning that relate to normal neurobiological changes rather than neurodegenerative disease processes (Harada, Natelson-Love, & Triebel, 2013; Salthouse, 2009; Salthouse, 2010; Whalley, Deary, Appleton, & Starr, 2004). Indeed, older adults make up a substantial portion of many neuropsychologists’ patient populations, and assessment for dementia is one of the most common referral questions (Rabin, Paolillo, & Barr, 2016; Sweet, Benson, Nelson, & Moberg, 2015). But assessing and understanding any patient, including an older adult patient, necessitates more than merely having knowledge of how brain pathology affects behavior; it requires recognition of the whole person, including their social and cultural context. Older adults face widespread discrimination and stereotyping, and this prejudice is frequently based on assumptions of cognitive decline in later life (Whitbourne & Sneed, 2004; WHO, 2015). These assumptions raise concern for the ability to objectively assess cognitive decline in seniors, as stereotypes may negatively impact objective performance through altered performance expectations, through increased evaluative pressure on an individual at risk of confirming negative stereotypes, or both. Yet while neuropsychologists are increasingly attending to non-neurological contributors to cognitive test performance such as social and cultural factors among many patients, cultural factors specifically related to older adults appear largely ignored, and more research on these factors is needed to inform clinical practice with older adults. To that end, the present study aimed to examine the effects of two psychosocial factors theorized to negatively impact older adults’ cognitive test performance: stereotype threat and diagnosis threat.
Stereotype Threat

In a pioneering paper exploring the impact of negative stereotypes about African Americans and intellectual functioning on academic achievement tests, Steele and Aronson (1995) coined the term *stereotype threat*. As originally conceptualized by the authors, stereotype threat represents the potential for negative performance impact when an individual is placed in circumstances in which they are at risk of confirming a negative stereotype about their group. This effect is theoretically unrelated to belief in the stereotype, but rather relates only to the relevance of a stereotype to the domain being assessed and a concern for one’s performance or behavior supporting the stereotype.

Across a series of four studies, the authors demonstrated that African American students in stereotype threat conditions underperformed on tests of verbal ability and academic achievement as compared to African American students in control conditions and White participants in either condition. Additionally, as compared to those in the other conditions, African American students in threat conditions showed greater levels of stereotype activation (filling in more stereotype-related words in a word completion task), stereotype distancing (such as endorsing lower levels of enjoyment in things stereotypically associated with African Americans, such as rap music), and self-handicapping (e.g., reporting getting less sleep the night before or rating the tests as more unfair). These stereotype threat effects were elicited through simple manipulations that represent routine aspects of intellectual and cognitive testing, such as by implying the tests were diagnostic (e.g., explaining that test results “may be helpful to you by familiarizing you with some of your strengths and weaknesses” and that the tasks represented a “genuine test of your verbal abilities and limitations so that we might better
understand the factors involved in both”) or even by simply asking participants to indicate their race on a demographic form prior to testing.

While research demonstrating the potential negative effects of racial tension on testing performance of African American students dates as far back as the Civil Rights era (e.g., Katz, Epps, Axelson, 1964; Katz, Roberts & Robinson, 1965), Steele and Aronson’s work introduced speculation that the phenomenon was something more than the side effects of a tumultuous era of racial clashing. Instead, they speculated that the driving factor was being at risk of confirming a negative stereotype about one’s group—a position in which anyone of any negatively stereotyped group could find themselves, regardless of whether they themselves agree or disagree with the stereotype. Subsequent research has pursued this line of reasoning by examining stereotype threat in additional groups. For example, Latino students in a stereotype threat condition in which a working memory task was framed as predictive of intellectual functioning performed more poorly as compared to Latino students in a control condition and as compared to White students in either condition (Schmader & Johns, 2003). Other research has demonstrated stereotype threat effects among French students of low socioeconomic status (SES) given verbal problems; such individuals were outperformed by high SES individuals when the task was presented as assessing intellectual ability—but not when the task was presented as simply assessing relationships between attention and word memory (Croizet & Claire, 1998). Spencer and Castano (2007) found comparable results with a sample of American students, and additionally found the lowest levels of reported confidence in task performance among low SES students whose lower income status was made salient in a stereotype threat condition.
Stereotype threat does not exclusively impact minority groups or those that have been stigmatized, but theoretically can affect an individual of any group for which stereotypes exist. For example, Spencer, Steele, and Quinn (1999) demonstrated stereotype threat effects on women’s performance on math tests (an academic domain with negative stereotypes regarding women’s capability). Across three studies, men outperformed women only in threat conditions, including a condition in which participants were told a math test typically showed gender differences (as compared to a condition in which participants were told there were no gender differences) and when no information regarding typical gender performance was mentioned (again, compared to a condition in which participants understood there to be no gender differences on the test at hand). This effect held for male and female students with a history of high math achievement as well as for those with more mid-range achievements. Among high achievers, the effect was observed on difficult but not easy tests, suggesting the threat may have been dependent upon recognizing poor performance and subsequent concern for confirming stereotypes (Spencer et al., 1999). Further research has suggested that this stereotype threat effect is contingent upon group identification—for example, stereotype threat effects on math performance among women might only be exhibited among women who report their gender as an important part of their identity (Schmader, 2002). Even White males, a group for whom few negative stereotypes pervade American culture, are not immune: evidence suggests that their own math performance is negatively impacted when they believe they are to be compared to stereotypically mathematically superior Asian test-takers (of unspecified gender), and that this effect is particularly
pronounced among White males who place a high importance on mathematics skills as part of their self-concept (Aronson et al., 1999).

In short, evidence suggests that performance on cognitive testing can be negatively impacted based on an individual’s perception of being at risk of confirming a negative view of him/herself or the group to which he/she belongs. In general, these groups refer to categories people find themselves in for their whole lives (e.g., racial/ethnic background or gender). But one widely stereotyped population represents a group into which everyone fortunate enough to live long enough will join: older adults.

**Stereotype Threat and Older Adults**

Older adults represent a frequently stigmatized group. Researchers suggest age-related stereotyping and ageism are not limited to Western cultures (Cuddy, Norton, & Fiske, 2005), and the World Health Organization speculates that ageism may be more prevalent than other more commonly recognized prejudices such as sexism and racism (WHO, 2015). (For a recent review of the literature on ageism, including so-called “positive” ageism, see Levy & Macdonald, 2016). A survey of Americans over the age of 60 and Canadians over 50 suggested that the vast majority of older adults (84% of Americans and 91% of Canadians) have experienced at least one incident of ageism. The most frequent such events were seemingly benign, such as hearing jokes or receiving birthday cards that make fun of older adults—though as the author observes, such experiences nevertheless may be considered ageist as they are based on age-related stereotypes. Perhaps of greater concern were the assumptions older adults reported facing: 30% of Americans reported at least one incident of being assumed to be deaf or hard of hearing, while 33% reported one or more incidents of facing assumptions that
they could not understand something due to their age. Forty-two percent reported a medical provider assumed their medical symptoms or illnesses were due to age, and as the author notes, “Some respondents indicated on their forms that they agreed that their ailments were caused by age, but in fact chronological age does not ‘cause’ anything” (Palmore, 2004, p. 44). Among a national sample of adults age 50 and older, 8.3% reported some worry about being judged for their age by a medical provider (Abdou, Fingerhut, Jackson, & Wheaton, 2016).

These experiences of ageism may come as no surprise after examining research on attitudes toward older adults. One meta-analysis suggested that when compared with attitudes toward younger adults, older adults are consistently viewed more negatively and are highly likely to be judged according to stereotypes—particularly in low context situations where few details about an individual are available (Kite, Stockdale, Whitley Jr., & Johnson, 2005). The content of such stereotypes regarding older adults frequently includes perceptions that they are lonely and have poor coping skills, that all older adults are alike, that they are dependent on others due to illness and frailty, and that they are cognitively impaired (Whitbourne & Sneed, 2004). Related to the assumption of cognitive impairment, the stereotype content model (Fiske, Cuddy, Glick, & Xu, 2002) has explored the assumed incompetence of older adults as part of a larger exploration of the mixed nature of stereotypes. According to the stereotype content model, stereotypes are generally classified as existing on two dimensions: competence and warmth. Most out-group stereotypes are associated with a mixture of positive and negative; that is, most fall into the broad domains of “envious stereotypes” (characterized by perceptions of a group as having high competence but low warmth), or “paternalistic stereotypes” (in
which a group is viewed as having low competence, but high warmth). Competence perceptions are theorized to be driven by perceptions of power such that those with higher power and social ranking are viewed as more competent, while warmth perceptions are driven by perceptions of competition (such that those not viewed as a competitive threat are rated as warmer). In this model, stereotypes of older adults are classified as paternalistic stereotypes characterized by ratings of high warmth but low competence (and thus, viewed as having low power or social rank and as additionally not being a competitive threat to others).

Given the pervasive nature of stereotypes regarding older adults’ functional and cognitive capacities, it comes as no surprise that researchers have explored the effects of stereotype threat among older adults. While some studies have explored threat effects related to stereotypes concerning older adults’ physical functioning (e.g., Swift, Lamont, & Abrams, 2012) or driving ability (e.g., Gaillard, Desmette, & Keller, 2011; Joanisse, Gagnon, & Voloaca, 2013), the majority of studies have explored the stereotype threat effects on older adults performance on tests of cognitive functioning. Memory represents the most frequently assessed domain. In these studies, threat is most commonly induced via instructions drawing attention to a task’s memory-evaluative nature (e.g., “You will be asked to remember a list of words. . .”), by indicating older adults’ performance will be compared to younger adults, by asking participants to read brief pseudo-newspaper articles presenting research that summarizes the cognitive/memory decline inherent in aging, or a combination of these manipulations. Older adults in these conditions are compared to either a neutral control condition or a stereotype-challenging condition—individuals in the latter example might, for example, read an article stating older adults
do not experience memory declines with age, or that the tests in the present study are age-
neutral. Although some studies include only older adults, others include wider age ranges
to allow comparison of effects between older adults and young or midlife.

Although some variability in the older adult stereotype literature exists, the
majority of published studies demonstrate at least a small but generally consistent
stereotype threat effect such that memory and/or general cognitive functioning is
dampened by stereotype threat manipulations. In one of the earliest applications of
stereotype threat research to older adults, Rahhal, Hasher, and Colcombe (2001) found
that on a task of accurately recalling whether previously learned statements were true or
false, older adults were outperformed by younger when the word memory was
emphasized in instructions (e.g., “You will be tested on your memory of this information
in phase two,” p. 700), but performed equivalently to younger adults when the word
memory was not used in instructions (e.g., “You will be tested on this information in
phase two,” p. 700). Other studies have suggested that young adults outperform older
adults on cognitive tasks regardless of whether or not memory is included in instructions
(consistent with evidence for normal changes in cognitive aging), but that differences in
performance are greater in threat conditions using the word. In one such study, younger
adults performed better than older adults across tests of immediate verbal recall,
immediate visuospatial recall, visuospatial span, and a combined memory and spatial
navigation task, and young adults were unaffected by different instructions. Among older
adults however, when tasks were presented as memory-based, performance was worse
compared to those for whom tasks were presented as tasks of orientation or described
vaguely as general cognition measures. This effect was especially pronounced when
memory self-efficacy (i.e., belief in one’s ability to effectively use and control one’s memory) was lower. Furthermore, memory-emphasized instructions were associated with lower performance expectations for older adults (Desrichard & Köpetz, 2005). In another study, participants in a threat condition involving explicit mention of memory evaluation, formal test booklets, listing of participant age at the top of each test page, and presence of a younger person in the room performed more poorly on a story memory free recall task than participants in a non-threat condition with no mention of memory testing, less formal looking test materials, no age indication on each page, and an older confederate present in the room—although cued recall and recognition memory for the same short stories did not demonstrate any threat effects (Kang & Chasteen, 2009).

Still, other studies have suggested that use of the word memory in instructions has no impact on older adults’ performance or perceived threat. Chasteen, Bhattacharyya, Horhota, Tam, and Hasher (2005) found that instructing participants to form an overall impression of a person based on behavioral descriptors (memory de-emphasized, non-threat condition) was associated with better performance among both older and younger adults as compared to those instructed to memorize the descriptors (memory emphasis, threat condition). Furthermore, older adults in both threat and non-threat conditions reported perceived stereotype threat on an explicit scale including questions such as “The experimenter expected me to do poorly because of my age.” Additionally, younger adult performance was better than that of older adults regardless of condition. Nevertheless, evidence across multiple studies suggests stereotype threat effects may negatively impact test performance of older adults who are aware of taking a memory-diagnostic test. Clinically, this has serious implications for neuropsychological evaluations of older
adults, who are almost certainly aware of the diagnostic nature of the tests they are taking.

Presenting participants with pseudo-newspaper articles regarding older adults and memory/cognitive functioning to induce (or eliminate) threat is also a common manipulation. For example, in one study younger (18-30 years) and older (62-84 years) participants either (a) read a fabricated newspaper article about declines in memory functioning with age (negative stereotype condition), (b) read a fabricated newspaper article about memory not changing with age (positive stereotype condition), or (c) read nothing (control condition). Older adults in the negative condition performed worse on a word list immediate recall measure as compared to other older adults in the positive and control conditions, and as compared to younger adults in all conditions (Hess, Auman, Colcombe, & Rahhal, 2003). In addition to free recall, recognition discriminability may also be impacted by such interventions. For example, Thomas and Dubois (2011) used a Deese-Roediger-McDermott paradigm in which participants studied 12 lists of 15 words each (with all words in a given list related to a specific semantic category). Participants were then either read a statement about research on declines in memory associated with aging and told the study was about memory (high-threat condition), or read a statement about research on language processing and told the study was about language functioning (low-threat condition). Participants then completed a yes/no recognition test for the list words and distractor words, and additionally rated their confidence in their judgments. Older adults in the high-threat condition had higher rates of endorsing category-related distractor words as compared to those in the low-threat condition and as compared to younger adults in either condition. They also had higher confidence in their (incorrect)
judgments as compared to the other conditions. Even when manipulations do not explicitly state that cognitive decline is assumed, but rather indicate awareness of the stereotype and inform participants that the research is intended to investigate the veracity of those assumptions (while also noting that older and younger people will participate), stereotype threat may be observed when compared to conditions where no mention is made of possible differences (Abrams et al., 2008; Abrams, Eller, & Bryant, 2006). Still, another study reported no effect on word list immediate recall when using manipulations based on presenting participants aged 60-75 with fabricated research with stereotype-consistent or stereotype-inconsistent information (Horton, Baker, Pearce, & Deakin, 2010)—although it should be noted that this study statistically controlled for variables theorized to moderate stereotype threat effects (e.g., age and investment in memory/physical functioning). In any event, given the wealth of articles in the popular press regarding aging and cognitive changes as well as services advertised to address such change (e.g., Lumosity), studies showing stereotype threat effects arising from reading “research” similar to popular press coverage of older adults’ cognitive decline suggests reason for concern.

Another frequent stereotype manipulation is indication of intent to compare performance of older adults to another group (typically younger adults in their 20s-30s). In one example, although all participants were told they were to complete a memory test, in the threat condition age differences were emphasized and participants understood such differences were being evaluated, whereas in the nonthreat condition participants were led to believe the test was age neutral, and that adults of any age perform similarly. Adults aged 60-70 performed more poorly on a word list immediate recall measure in the
threat condition as compared to the nonthreat condition (Hess, Hinson, & Hodges, 2009). Comparable results were found in a slightly older French sample using common cognitive screening tools as the dependent measures—being told both young and old adults would participate in a study of memory (as opposed to being told both would participate but that tests are not affected by age) was associated with poorer performance (Mazerolle et al., 2016). In another study, older adults’ performance on working memory and cued recall performance was poorer in a threat condition in which they were told both younger and older adults were participating as compared to (a) older adults in a reduced-threat condition who were told the tests would not show age difference, and (b) young adults in either condition (Mazerolle, Régner, Morisset, Rigalleau, & Huguet, 2012). When a time limit for task performance was involved (thereby increasing task difficulty as compared to a group given no time limit), general mention of age differences (e.g., “One goal of this study is to examine age differences in memory ability . . . Younger adults typically do much better than older adults on this task,” p. 483) was sufficient to dampen performance on a 50-word list recognition measure as compared to those who were provided no information regarding comparative intentions (Hess, Emery, & Queen, 2009). Threat effects may not even be dependent on assumed comparison to young adults. When same-age participants believed different age groups were to be compared and that they were at the older end of the participant age range of 40-70 (rather than at the young end of an age range of 60-90), they exhibited poorer performance on a memory task and/or a general cognitive task. Specific deficits were dependent on whether participants had read an article reporting age is associated with memory deficits or one suggesting age is associated with global cognitive dysfunction (Haslam et al., 2012).
**Moderators and mediators of age-based stereotype threat.** In addition to different methodologies for inducing threat, extant studies have explored different possible moderators of stereotype threat among older adults. Consistent with stereotype threat’s prediction that threat is operational among those who place high value on the domain of interest (Steele & Aronson, 1995), several investigations of stereotype threat and older adults suggest that placing a higher value on memory functioning relates with greater threat susceptibility (Hess et al., 2003; Joanisse et al., 2013). Other studies suggest that memory self-efficacy may influence or moderate stereotype threat effects, finding stronger threat effects among those with lower memory self-efficacy (Desrichard & Köpetz, 2005). Baseline memory self-efficacy may moderate threat effects such that those with lower self-efficacy perform more poorly under threat conditions and those with higher self-efficacy receive a performance boost (Schlemmer & Desrichard, 2017), though it should be noted that effects of baseline memory self-efficacy are not consistently observed (Bouazzaoui et al., 2016).

Age itself may be a moderating factor. For example, in a study of participants ranging from 24 to 86 years, fabricated news articles discussing inevitable age-related memory declines for older adults actually resulted in “stereotype lift” for adults in midlife, whose performance on memory tasks saw an improvement (ostensibly through downward social comparison with a group of some relevance, as midlife adults move toward older adulthood; Hess & Hinson, 2006). Similarly, other studies have observed threat effects in those transitioning to older adulthood (ages 60-70), but not those further into that phase of life (ages 71-82; Hess, Hinson, & Hodges, 2009). Subjective age may also play a role—for example, greater identification with older adult groups has been
shown to relate to worse performance on free recall measures (Kang & Chasteen, 2009). This is consistent with research on self-stereotyping effects in which priming older adult stereotypes among those for whom such stereotypes are on the cusp of becoming self-relevant might influence performance (O’Brien & Hummert, 2006). Research on these effects also emphasizes the importance of the relevance of aging stereotypes (Levy, 2009), recognizing that stereotypes become self-relevant in stages based on both social factors (as others begin to view one as an older adult) and individual factors (as one begins to view oneself as an older adult; Levy, 2003). While self-stereotyping represents a different theoretical perspective on the impact of age-related stereotypes on behavior and cognition, its basic underlying assumptions (e.g., that older adults are impacted by stereotypes internalized over the lifespan) hold true for an age-based stereotype threat perspective as well. This may suggest a need to examine subjective age and age-related stereotype self-relevance in stereotype threat effects. Still, as some researchers observe, those who are viewed by others as older adults often don’t see themselves that way—which could have important implications for the specific type of stereotype threat they experience (i.e., one that reflects concerns related to self-image more so than group image; Barber, 2017).

In the same vein of demographic factors, education may also moderate stereotype threat effects—although studies present conflicting information as to precisely how. For example, Andreoletti and Lachman (2004) suggest that for highly educated adults across the lifespan, exposure to counterstereotype information—that is, being told a memory test does not show age-related differences—improves performance, whereas for those with less than a college degree, exposure to either counterstereotype or stereotype-consistent
information worsened performance relative to control. However, a more recent study has suggested that higher education (a variable considered by the authors as reflective of the value one places on cognitive functioning) is actually associated with greater susceptibility to threat (Hess, Hinson, & Hodges, 2009).

Social variables may also play a role in threat susceptibility: a group of researchers led by Abrams reported positive intergenerational contact reduced or eliminated stereotype threat effects, even when that contact was merely imagined (Abrams et al., 2006; Abrams et al., 2008). More broadly, stigma consciousness (Hess, Hinson, & Hodges, 2009) or perceived threat as measured by explicit threat perception measures (Kang & Chasteen, 2009) may relate to threat induction susceptibility and/or worsened performance.

Lastly, difficulty of the task may moderate stereotype threat, as evidenced by previous literature showing no threat effects for high achieving women on easy math problems, but evidence of stereotype threat on more difficult ones (Spencer et al., 1999). There has been evidence of this within the older adult literature as well—for example, Hess, Emery, and Queen (2009) found threat effects only among those for whom a memory recognition task had a strict time deadline to respond, but not among those who had unlimited time to make a decision.

These moderation effects raise the question of what mechanisms drive stereotype threat effects. Unfortunately, evidence remains unclear as to precisely which variables mediate threat’s effect on functioning. Steele and Aronson’s (1995) original examination of stereotype threat conceptualized it as a “self-evaluative threat” with numerous possible mechanisms, including over-arousal, redirection of attentional resources to stereotype-
related worry, increased cautiousness in test performance, or withdrawn effort. In a recent review, Spencer, Logel, and Davies (2016) summarized the major mediational mechanisms considered in stereotype threat literature more broadly. These include the increased pressure to perform well, which may result in several possible performance-worsening outcomes such as a tendency to rely on automatic responding, overconsumption of working memory resources, and having attention drawn to typically automatic processes. Another reported mechanism includes use of strategies to attempt to protect the self from risk of confirming a stereotype, such as putting forth less effort/reducing their self-expectations, endorsing other explanatory mechanisms for poor performance (e.g., poor sleep), not practicing, or not attempting as many items (Spencer et al., 2016). Of note, however, some researchers propose stereotype threat may not be a one-size-fits-all construct, and that instead different types of threat exist, potentially driven by different mediational mechanisms (Barber, 2017; Shapiro & Neuberg, 2007). Indeed, there is evidence to suggest that threat may be mediated by different variables for older adults than it is for young adults (Popham & Hess, 2015).

However, conclusions on what mediational mechanisms drive stereotype threat in older adults are no easier to draw than for young adults. Even evidence for explicitly reported perceived threat has been inconsistent, with some studies suggesting this may mediate effects (e.g., Chasteen et al., 2005) and others finding no support for this (e.g., Barber, Mather, & Gatz, 2015). Given the consideration of evaluation apprehension as a major underlying factor in stereotype threat, one might expect anxiety to be identified as a consistent mediator. However, within the literature on stereotype threat in older adults, there is little evidence to support increased anxiety as an explanatory mechanism:
multiple studies have found no evidence of a mediational mechanism of anxiety (Barber et al., 2015; Hess & Hinson, 2006; Hess, Hinson, & Hodges, 2009). Abrams and colleagues, however, have found a mediated moderation mechanism in the form of reduced anxiety via positive intergenerational contact, which in turn moderates threat effects (Abrams et al., 2006; Abrams et al., 2008).

Evidence for depletion of cognitive resources as an explanatory mechanism of stereotype threat effects has also been inconsistent. In support of this possibility, one study’s threat effects observed on a recognition memory test with limited time to respond (when such effects were not observed with unlimited time to respond) may reflect reduced speed of processing. Results from this same study suggesting that those in the threat condition were likelier to endorse a feeling of knowing versus consciously remembering items may also reflect inefficiency of processing (Hess, Emery, & Queen, 2009). This is consistent with research suggesting stereotype threat may cause older adults to rely more on automatic response styles, leading to more incorrect answers on more difficult tasks (Mazerolle et al., 2012). This same study suggested some evidence for reduction in working memory as a partial explanatory mechanism for threat effects in older adults (Mazerolle et al., 2012), comparable to mechanisms outlined in other populations (Schmader, Johns, & Forbes, 2008). However, other research has suggested that reduced working memory may not drive stereotype threat effects in older adults (Hess, Hinson, & Hodges, 2009; Popham & Hess, 2015), and mechanisms may differ between stereotype threat effects in younger adults versus those in older adults (Barber, 2017).
Regulatory focus (Higgins, 1997) has been proposed both as a mediator and moderator of threat effects. In its simplest form, this theory considers behavior regulation as being either promotion focused (seeking gains) or prevention focused (minimizing losses). From this framework, stereotype threat is considered to induce an increased focus on preventing losses (to prevent confirming negative stereotypes of poor performance) as opposed to seeking gains, resulting in a more conservative test-taking approach that may actually worsen performance. Gaillard et al. (2011) suggest that from a developmental perspective, loss prevention may already be more predominant for older adults, and that their performance may therefore be particularly impacted by stereotype threat. This speculation was supported in the authors’ own research, in which threat effects were observed when older adult participants were encouraged to try to avoid mistakes, but not when participants were told to try to get many items right. Interestingly, a regulatory focus explanation for age-based stereotype threat effects suggests that older adults may actually benefit from stereotype threat in situations in which a prevention-focused approach is warranted. For example, Wong and Gallo (2016) conducted a modified replication of Thomas and Dubois’s (2011) work using similar priming procedures and the same recognition paradigm, exposing young and old participants to 12 lists of 15 words. However, prior to recognition memory testing that included lures semantically related to list words, the researchers warned all participants about the risk of false positive responding inherent in the task (in short, encouraging more conservative responding). Older adults under threat conditions endorsed fewer false positives relative to older adults in the control condition (opposite findings from Thomas & Dubois, 2011), a result not seen in the younger adult group. In short, stereotype threat appeared to better
enable older participants to use loss-avoidant/prevention-focused strategies, resulting in fewer false positives (though notably, they also endorsed fewer true positives). In another study, stereotype threat detrimentally affected performance on tasks where consequences encouraged gains-based responding, but not when consequences encouraged loss-avoidance (Barber et al., 2015). While these studies support a moderation effect, other studies have also reported stereotype threat inducing performance that sacrificed speed of performance for accuracy, suggesting a mediational mechanism whereby threat induced a loss-prevention strategy that resulted in overall worse performance (albeit with fewer errors; Popham & Hess, 2015). Indeed, regulatory focus has recently been proposed as potentially the primary mechanism underlying age-based stereotype threat specifically (in contrast to other types of threat, such as gender and ethnicity; Barber, 2017).

Lastly, lowered self-expectations have been proposed to play a role in threat effects (Steele & Aronson, 1995). For older adults, stereotype information may impact aging concerns and perceptions of having control over one’s memory abilities—which in turn may drive performance on memory tasks for better or worse (Hess & Hinson, 2006). Memory self-efficacy, along with subjective memory complaints, have both found to be mediators of threat effects on memory test performance (Bouazzaoui et al., 2016). Reduced expectations for cognitive performance, whether for older adults as a group or for the individual in particular, may also explain observed threat effects (Desrichard & Köpetz, 2005; Hess, Hinson, & Hodges, 2009).

**Age-based stereotype threat summary.** Clearly, there is some heterogeneity of observed effects among age-based stereotype threat research. A recent meta-analysis (Lamont, Swift, & Abrams, 2015) including both published and unpublished studies cited
an overall effect size of $d = .28$ (a small to medium effect), but noted that specific study considerations resulted in widely varying effect sizes. Overall, stronger effects were found in studies that evaluated multiple cognitive domains ($d = .68$) rather than memory exclusively, and in studies that used “stereotype-based” rather than “fact-based” manipulations (e.g., those that simply invoked stereotypes by mentioning a planned age comparison tended to be more effective than studies that presented participants with “research studies” about changes in aging). Results also suggested strongest effects with dependent variables assessed immediately after threat induction, with progressively weaker effects at each administration position afterward. While some specific studies suggested stereotype threat might be more operational among those in the earlier phase of older adulthood (e.g., Hess & Hinson, 2006), meta-analytic results suggested that age did not account for differences in effect sizes.

Regardless of variability in the strength of effects or the precise domains affected by stereotype threat, what is clear from research in this vein is that cognitive test performance has the potential to be impacted by salient, relevant stereotypes for examinees. In line with stereotype threat’s social psychological roots, this research has primarily focused on social identities (e.g., race, gender, age) as the stereotyped domain. More recently, neuropsychological researchers have begun to explore similar effects not based on demographic identity, but neurological diagnoses.

**Diagnosis Threat**

Inspired by research on stereotype threat and the broader concept of psychosocial influences on neuropsychological test performance, Suhr and Gunstad (2002) proposed the concept of “diagnosis threat” among persons with a history of mild traumatic brain
injury (mTBI)—an injury of minor severity, typically associated with complete recovery within weeks to no more than 3 months in nonlitigating samples (Belanger, Curtiss, Demery, Lebowitz, & Vanderploeg, 2005; Schretlen & Shapiro, 2003). According to the authors, diagnosis threat describes the phenomenon whereby those with mTBI might underperform on cognitive testing when their injury and associated symptoms were made salient. In their experimental design, Suhr and Gunstad (2002) recruited students with a history of mTBI reported in a separate, unrelated data collection process. The authors randomly assigned participants to one of two conditions with varying instructions, such that the mTBI history (and associated deficits) were either made salient or ignored altogether. Instructions for the diagnosis threat condition were as follows (underlined segments indicate sections included only in the diagnosis threat condition, whereas non-underlined segments were included in both diagnosis threat and control conditions):

You have been invited to participate in this study because of your responses to one of the questionnaires included in the mass screening at the beginning of the quarter. Your responses indicated a history of head injury/concussion. A growing number of neuropsychological studies find that many individuals with head injuries/concussions show cognitive deficits on neuropsychological tests. Deficits in areas such as attention, memory, and speed of information processing are common – though other deficits sometimes emerge. This study examines the role that head injury may play in these cognitive areas to better understand the nature of the disorder. When the experimenter returns to the room, s/he will ask you to complete a brief collection of common neuropsychological tests. These tests will assess skills such as attention, memory, speed of information processing, problem
solving skills, etc. Some of the tests are easy, some are more difficult. Please give your best effort. Questions about individual tests will be answered following the testing. (pp. 450-451)

After reading these instructions, participants completed a brief neuropsychological battery including several routinely used neuropsychological tests spanning multiple cognitive domains. They also completed a self-report measure assessing effort. For those in the diagnosis threat group, performance was lower on tests of immediate recall, delayed recall (both verbal and visuospatial), and “intellectual functioning” (including a test of general knowledge and a test of visuo-constructional ability). Attention/working memory, psychomotor speed, and verbal fluency were not affected. Self-rated effort was lower in the threat group and correlated with some of the affected tests, leading the authors to conclude that reduced effort may have driven poorer performance (Suhr & Gunstad, 2002).

However, a follow-up study (Suhr & Gunstad, 2005) specifically designed to investigate this effect suggested otherwise. This time, the authors explicitly assessed effort via performance validity testing—measures that are objectively easy, but may appear difficult and therefore elicit poor performance by those who commit suboptimal effort. Additionally, the authors attempted to assess the relationship of psychological variables to diagnosis threat by assessing depressive symptoms prior to testing and anxiety symptoms after threat induction. The authors again found significant effects of condition on cognitive performance (albeit on some different tests this time, with effects now found on tests of attention and working memory, such as a digit span task). Notably, while 12% of participants in the control condition performed in a range classifiable as
impaired based on published test norms for at least one test, in the diagnosis threat condition, 46% of participants performed in such a range. However, in this study, conditions did not differ on self-reported or objectively measured effort. Additionally, neither anxiety nor depression played any role in effects observed.

Follow-up studies of diagnosis threat among mTBI populations have been less consistent in replicating cognitive effects of diagnosis threat, but have raised suspicions for the impact of diagnosis threat on relevant psychological variables. For example, Trontel, Hall, Ashendorf, and O’Connor (2013) used a design nearly identical to Suhr and Gunstad’s 2005 work, with the notable addition of a measure of academic self-efficacy. They found cognitive effects of diagnosis threat only on a test of general knowledge (an effect also found in Suhr & Gunstad, 2002), and no differences between diagnosis threat and neutral conditions on either self-reported effort or an objective effort measure (consistent with Suhr & Gunstad, 2005, but inconsistent with the 2002 study). However, they also identified lower levels of reported academic self-efficacy among students with a history of mTBI in the diagnosis threat condition as compared to those in the neutral condition. Relatedly, among a non-student sample diagnosis threat may negatively impact memory self-efficacy (Kit, Mateer, Tuokko, & Spencer-Rodgers, 2014)—although this study must be interpreted cautiously in light of their inclusion of some participants with suspected moderate TBI.

In addition to self-efficacy, diagnosis threat may influence symptom reporting. Ozen and Fernandes (2011) explored diagnosis threat versus neutral instructions with the inclusion of control participants with no history of mTBI. Similar to other replication attempts, their findings of the effect of diagnosis threat on cognitive functioning were not
as striking as Suhr and Gunstad’s. In this study, a main effect of diagnosis was observed such that control participants outperformed those with mTBI on only a digit span task. Although the interaction between diagnosis and threat induction was nonsignificant, planned comparisons were nevertheless examined and suggested that controls outperformed mTBI participants in the diagnosis threat condition. Outside of this planned comparison of a nonsignificant interaction, there was no substantial evidence of impact of diagnosis threat induction on cognitive functioning—although it should be noted this study used a more mildly worded threat induction and additionally, did not include measures of delayed recall. A more striking finding in this study was subjective symptom reporting: those with a history of mTBI in the diagnosis threat condition endorsed more complaints of everyday lapses in attention as compared to mTBI participants in the neutral condition and as compared to those without a history of mTBI. Additionally, participants with a history of mTBI in the diagnosis threat condition reported more everyday memory complaints than those with no history of mTBI—whereas the two participant groups were not significantly different from each other in the neutral condition. The results of this study suggest that diagnosis threat may be a significant concern not only for its potential impact on cognitive testing results, but also for its inflating influence on patient self-report of symptoms. However, other studies have failed to note differences between diagnosis threat conditions and a neutral control condition among participants with a history of mTBI in terms self-reported somatic, cognitive, or affective symptoms (Blaine, Sullivan, & Edmen, 2013; Carter-Allison, Potter, & Rimes, 2016).
Taken together, while initial studies of diagnosis threat demonstrated effects of making one’s diagnostic category (and presumed deficits) salient on cognitive functioning, subsequent studies have had better luck identifying effects on psychological variables such as self-efficacy and self-reported functioning. Some authors speculate that a lack of observed cognitive effects may be due to a failure to ascertain participants’ level of identification with the injury; for example, among male participants, a moderately strong effect of diagnosis threat on an arithmetic working memory task was found when analyses controlled for injury identification (as measured via questions such as “Being an individual who has sustained a concussion is an important reflection of who I am”; Pavawalla, Salazar, Cimino, Belanger, & Vanderploeg, 2013). Similar to the more extensively studied stereotype threat, literature on diagnosis threat has yet to consistently identify mediational mechanisms to explain the effect’s operation. As discussed above, proposed mechanisms include reduced effort, anxiety/negative affect, and lowered self-efficacy—however, none of these has been firmly established.

**Concepts related to diagnosis threat.** The concept of non-neurological contributors to cognitive and somatic functioning after neurological injury is not a new one. Specific to mTBI, researchers have explored psychosocial variables, including those that contribute to postconcussion syndrome (PCS)—a combination of fairly nonspecific symptoms (e.g., headache, fatigue, memory lapses) often experienced after mTBI, but also commonly experienced by non-injured persons. (When these symptoms persist outside of the window in which residual effects of the injury would be expected, they may instead be termed “postconcussive disorder” [PCD; McCrea, Janecek, Powell, & Hammeke, 2014].) In one influential study, control participants with no history of or
exposure to head injury completed a checklist of symptoms for their current functioning before being asked to imagine a car accident associated with a temporary loss of consciousness and hospitalization. Next, participants were asked to rate what symptoms they might expect to have six months post-accident. Responses of these participants with no history of head injury clustered into a reliable, PCS-consistent symptom set endorsed as current symptoms by head-injured participants—leading the authors to identify expectation as the etiology of somatic and cognitive dysfunction. Interestingly, this study also found that when participants with head-injury were asked to rate their pre-injury symptoms, they reported significantly fewer symptoms than non-injured participants reported as their own current functioning (Mittenberg, DiGuilio, Perrin, & Bass, 1992). While some studies have replicated this disproportionate symptom reporting among those with head injury (Ferguson, Mittenberg, Barone, & Schneider, 1999; Iverson, Lange, Brooks, & Rennison, 2010), other researchers have found comparable reporting biases among those with head injury, depression, and tension headaches. They therefore advocate for a broader descriptor in the form of the “good old days” bias, “which suggests that any negative life event becomes the salient landmark for viewing current state as a negative change from the past” (Gunstad & Suhr, 2001, p. 330). Both “expectation as etiology” and “good old days” biases relate closely to the general concept of a nocebo effect—the antithesis of the more well known placebo effect—in which anticipation of negative changes drives such changes (Hahn, 1997).

Regardless of the specific terminology used or precise mechanism defined, these closely related concepts characterize a process by which concern about neurological diagnosis becomes disproportionate to observed cognitive dysfunction. Returning to
consideration of an older adult population, a related notion has recently been explored with a neurological diagnosis most commonly associated with a geriatric population: dementia.

**Dementia Worry**

Cutler and Hodgson (1996) initially proposed the concept of “anticipatory dementia” as a fear that normal changes in memory functioning may represent early warning signs of dementia processes. In the wake of further literature surrounding concern regarding the development of dementia, Kessler, Bowen, Baer, Frölich, and Wahl (2012) proposed the broader concept of dementia worry, “an emotional response to the perceived threat of developing dementia, independent of age and cognitive status,” including “an overlap of affective components (e.g., fear) as well as more cognitive components (e.g., associations, thoughts, images)” (p. 277). Dementia worry includes but is not limited to fear surrounding AD, and represents a spectrum of response intensity ranging from the level of a passing concern to the level of an outright phobia (Kessler et al., 2012).

Of note, while dementia worry includes affective and cognitive responses to the possibility of developing dementia, it is *not* implied that this relates in any way to cognitive decline or even objective risk. Indeed, for some proportion of the population, concern about developing AD or other dementias may trump objectively reported risk entirely. For example, Linnenbringer, Roberts, Hiraki, Cupples, and Green (2010) provided participants who had a first-degree relative with AD with a lifetime percentage risk of developing the disease themselves based on demographic factors and Apolipoprotein E (*APOE*) genotype—a genetic predictor of AD risk. Six weeks after this,
participants were contacted and asked to (a) recall the risk estimate they were given, and (b) state their perceived risk as measured by the question, “On a scale of 0-100%, what do you believe your chances are of developing Alzheimer’s disease sometime in your life?” Of the 158 participants who accurately recalled their risk estimate, approximately half provided a perceived risk within five percentage points of that range. However, one-third of participants who accurately recalled the risk estimate they were given nevertheless perceived their risk as more than five percentage points higher—and a surprising 23% of those who accurately recalled their risk estimate nevertheless ranked their perceived risk as more than 20 percentage points higher. Thus, for at least some participants, perceived risk felt higher than what objective estimates suggested—which may reflect a dementia worry mechanism at work.

Similarly, dementia worry represents more than mere realistic concern for dementia based on objective cognitive decline. Dementia worry has been repeatedly linked with subjective memory concerns; multiple studies have shown an association of poorer self-appraisal of memory with higher levels of dementia worry (Cutler & Hodgson, 1996; Cutler & Hodgson, 2001; Werner, 2002). This relationship has been evident both in cross-sectional studies of subjective appraisal of current memory and change in memory (Cutler, 2015) as well as in longitudinal analysis of the memory appraisal (Cutler & Brägaru, 2015). On the other hand, evidence for links between objective memory performance and dementia worry is less consistent (Kessler, Südhof, & Frölich, 2014; Kinzer & Suhr, 2015), and may be dependent on personal experience with dementia in a genetic or nongenetic relation (Suhr & Kinkela, 2007). Taken together, these results suggest that dementia worry represents something distinct from awareness
of true cognitive deficits that may be real warning signs of dementia. These results also indicate that patients presenting for costly, stressful, and resource-intensive neuropsychological testing due to perceived cognitive changes may be seeking such services unnecessarily.

Although one might expect dementia worry to rise with increased age—one of the strongest contributors to increased risk of AD and other dementias—research on this relationship is conflicting (and sometimes counterintuitive). Using a middle-aged sample of persons ranging from 40 to 60, Cutler and Hodgson (1996) showed that increased age was associated with lower levels of dementia worry. Data using a nationally representative sample with participants grouped into four age groups (50s, 60s, 70s, 80+) also suggested an overall decrease in dementia worry with increasing age (Cutler & Brăgaru, 2015). Another study using participants aged 50 through 85 suggested the same relationship for those who knew someone (related or unrelated) with AD, but possibly the opposite relationship for those with no Alzheimer’s experience—although the correlation between age and dementia worry in this group was nonsignificant (Suhr & Kinkela, 2007). A more recent study also divided respondents ranging from 55 to 90 years into groups based on dementia exposure, yet found no association between age and dementia worry for any groups (Kinzer & Suhr, 2015). Another study offers further evidence for no relationship, at least when comparing young-old (65-74) and old-old respondents (75-91; French, Floyd, Wilkins, & Osato, 2012). However, the bulk of current age-related dementia worry data is derived from cross-sectional research; it is therefore difficult to discern whether any age differences are due to age-related change or merely cohort differences.
Additionally, existing dementia worry research is primarily correlational in nature, with limited ability to differentiate between cause and effect in concern for developing dementia—for example, dementia worry may lead to depressive symptoms due to reflection on an unpleasant imagined future, or it may be a consequence of depression and associated negative ruminations (e.g., Suhr & Kinkela, 2007). However, one experimental study demonstrated that priming of negative age-related stereotypes increased levels of self-reported dementia worry as compared to those in a control condition and those who were primed with higher proportions of positive aging stereotypes. This effect was strengthened when controlling for the self-relevance of aging stereotypes as measured by the difference between chronological age and opinion on when old age begins (Molden & Maxfield, 2017). This suggests that as with cognitive functioning, concern for developing dementia can be influenced by aging stereotypes.

But based on existing literature concerning diagnosis threat and mTBI as well as the significant overlap between age-related stereotypes and assumptions of “senility” or dementia among older adults, the question is raised: Are observed effects on older adults’ cognitive functioning and self-reported cognitive and affective symptoms related to age-based stereotype threat or to dementia-related diagnosis threat? Or do each of these constructs produce equivalent results?

**The Present Study**

As Suhr and Gunstad (2002) noted, “It is important to remember that neuropsychological tests assess behavior and are not direct measures of brain function” (p. 448). While this was discussed in the context of mTBI patients, it is equally applicable across neuropsychological disorders and populations—including dementia and older
adults. In terms of social context, mTBI and dementia have a surprising amount in
common. Both concussion (generally used synonymously with mTBI) and AD (the most
common cause of dementia) have received enormous media attention in recent years.
Both have also been considered in major motion pictures. For example, the film
*Concussion* documented the experience of neuropathologist Dr. Bennet Omalu
identifying Chronic Traumatic Encephalopathy, a disorder hypothesized to result in
dementia via repeated concussive and subconcussive blows as is often the case in
organized sports such as football and soccer. The book-adapted-to-film *Still Alice*
presented a fictionalized account of Alice, a high achieving cognitive psychologist whose
cognitive functioning and functional independence were gradually stripped from her due
to early onset AD. As public awareness of both conditions increases, assumptions,
 misinformation, and stereotypes about the disorders can be expected to do the same. Yet
despite the potential similarity in terms of the social psychological variables that may
influence objective and subjective cognitive symptoms in those with a history of mTBI
and older adults at risk of developing dementia, existing published research has not
explored these similarities. Although a literature search revealed one unpublished
doctoral dissertation (Page, 2012) exploring effects of stereotype threat and diagnosis
threat related to dementia among older adults, the manipulations used in that project had
some overlap between experimental conditions (e.g., emphasizing aging as well as
dementia in the diagnosis threat group), making it difficult to differentiate between age-
based stereotype effects and dementia-related diagnosis threat effects. Additionally, while
results suggested an age-based stereotype threat effect on a recognition memory task and
an overall measure of cognitive functioning among participants who were not suspicious
of the manipulation, the project did not include consideration on how subjective variables such as self-reported cognition, dementia-specific anxiety, perceptions of stereotype threat, and participant effort may have been influenced by different threat manipulations.

The goal of the present research was to address this gap in the peer-reviewed research literature by exploring both stereotype threat and dementia-related diagnosis threat among older adults. As discussed above, stereotypes of older adults are almost inextricably associated with assumptions of cognitive decline or “senility.” This suggests that for older adults, stereotype threat effects may in fact be diagnosis threat effects. Alternatively, diagnosis threat may be something that is operational in a clinical setting (i.e., for older adults seeking diagnostic neuropsychological testing for cognitive concerns), whereas stereotype threat may reflect more of an everyday impact of daily, pervasive ageism—and may be more easily assessed in a lab, where participants are not necessarily expecting to be diagnosed with AD. If diagnosis threat (or stereotype threat) is observable on measures routinely used to diagnose Mild Cognitive Impairment (MCI), AD, or other syndromes related to cognitive decline among older adults, this is information neuropsychologists need to be aware of in clinical practice in order to find ways of mitigating this risk and ensuring optimal performance of patients. However, age-based stereotype threat has not received much attention in the field of neuropsychology, perhaps for two primary reasons: (a) existing stereotype threat studies of older adults are more frequently presented in gerontology/geropsychology/social psychology journals despite having obvious application and clinical relevance for clinical neuropsychologists, and (b) these studies almost exclusively use assessments not in routine use in clinical neuropsychology practice (including experimenter-created measures), and therefore may
not be reliably or validly assessing the same constructs neuropsychologists measure to diagnose dementia. Unlike more frequently used clinical measures, which generally have been evaluated for construct validity, experimenter-created measures have less empirical evidence to support that they measure the domain intended to be assessed. They may also vary in difficulty, which could affect estimates of how strongly threat affects performance—for example, being given two minutes to study a written word list before being asked to recall it may be easier than being read a word list and asked to recall it. At the very least, outcome measures outside of those used in the field of clinical neuropsychology may be presumed to have more restricted generalizability to clinical practice.

Finally, even if diagnosis threat and/or stereotype threat do not negatively impact cognitive functioning, they may nevertheless impact perceived functioning—as evidenced by diagnosis threat’s increase of self-reported memory and attentional problems in mTBI participants (e.g., Ozen & Fernandes, 2011), or by exposure to age-related stereotype information inflating concerns about developing dementia (Molden & Maxfield, 2017). Inflated concerns about developing dementia, increased subjective experience of symptoms, and/or misattribution of common cognitive symptoms (e.g., briefly misplacing something) to disease processes all carry risks of general distress, but also increase the chances of individuals seeking unnecessary cognitive testing. Additionally, the diagnostic criteria for MCI require subjective cognitive impairment (which can be reported by self or other) and impairment in at least one cognitive domain (Winblad et al., 2004). If literature on stereotype threat and diagnosis threat are to be believed, the impact of one or both of the constructs on (a) self-reported symptoms, and
(b) cognitive functioning in a threatening environment such as a neuropsychological evaluation, could easily contribute to the misdiagnosis of a cognitively intact older adult as having MCI.

The goal of the present research was to elucidate how stereotype threat and diagnosis threat might impact older adults’ cognitive functioning, self-reported cognitive complaints, and dementia worry. To do so, this study used a between-subjects design with three conditions: stereotype threat, diagnosis threat, and neutral control. Hypotheses were as follows (measures listed will be described in greater detail in the Method section):

**Cognitive variables.** Broadly, I hypothesized that cognitive outcome measures will be poorer among participants in the experimental conditions (stereotype threat induction and diagnosis threat induction) as compared to those in the control condition. I expected performance in the stereotype threat and diagnosis threat groups to be generally equivalent.

**Hypothesis 1.** I predicted poorer performance on tests of memory functioning among those in the stereotype threat and diagnosis threat conditions as compared to those in the control condition. Outcome measures of interest for this hypothesis included learning and immediate recall as measured by the California Verbal Learning Test Second Edition (CVLT-II) Trials 1-5 Total Recall raw score, delayed recall as measured by the CVLT-II Long-Delay Free Recall raw score, and recognition memory as measured by performance on the CVLT-II Long-Delay Yes/No Recognition trial.

**Hypothesis 2.** I predicted poorer performance on tests of auditory attention and working memory among those in the stereotype threat and diagnosis threat conditions as
compared to those in the control condition. Outcome measures of interest for this hypothesis included raw scores on the Wechsler Adult Intelligence Scale Fourth Edition (WAIS-IV) Digit Span forward and backward trials.

**Hypothesis 3.** I predicted poorer performance on tests of visuospatial attention and processing speed among those in the stereotype threat and diagnosis threat conditions as compared to those in the control condition. Outcome measures of interest for this hypothesis included time to complete the Delis-Kaplan Executive Function System (D-KEFS) Trail-Making Test (TMT) Conditions 2 (Number Sequencing) and 3 (Letter Sequencing), and time to complete D-KEFS Color-Word Interference Test (CWIT) Conditions 1 (Color Naming) and 2 (Word Reading).

**Hypothesis 4.** I predicted poorer performance on tests of executive functioning among those in the stereotype threat and diagnosis threat conditions as compared to those in the control condition. Outcome measures of interest for this hypothesis included D-KEFS TMT Condition 4 (Switching) and D-KEFS CWIT Condition 3 (Inhibition).

**Hypothesis 5.** I predicted different mediational mechanisms for cognitive impact of experimental condition such that for those in the stereotype threat condition, perceived stereotype threat would mediate cognitive performance, whereas for those in the diagnosis threat condition, perceived risk of AD would mediate cognitive performance.

**Self-report variables.** While performance among experimental groups was predicted to be roughly equivalent on cognitive measures (i.e., both groups performing worse than control), I anticipated the experimental manipulation groups may differ both from each other and from control on self-report variables.
Hypothesis 6. I predicted that self-rated change in cognitive functioning as measured by the short form of the Everyday Cognition scale (ECog-12) would differ across groups such that those in the diagnosis threat condition would report the greatest changes in cognitive functioning, those in the control condition would report the fewest, and those in the stereotype threat condition would rate themselves somewhere in between.

Hypothesis 7. I predicted that dementia worry as measured by the Dementia Worry Scale (DWS) would be highest in the diagnosis threat condition, lowest in the control condition, and somewhere in between among those in the stereotype threat condition.

Hypothesis 8. I predicted that perceived stereotype threat as measured by a modified version of Chasteen et al.’s (2005) questionnaire would be highest among those in the stereotype threat condition, lowest among those in the control condition, and fall in between the two for those in the diagnosis threat condition.

Hypothesis 9. I predicted that self-rated effort as measured by an item from a questionnaire used by Suhr and Gunstad (2002, 2005) would be highest among those in the stereotype threat condition, lowest among those in the diagnosis threat condition, and fall in between the two for those in the control condition.
CHAPTER II

METHOD

All methods were approved by the Institutional Review Board prior to participant contact and data collection.

Participant Characteristics

Eligibility and exclusion criteria. Eligible participants were those who were age 60 or older and who agreed to participate in exchange for $15. In addition, participants were required to have no evidence of cognitive impairment. For the present study, cognitive impairment was defined as performance two standard deviations below a demographics-based mean on more than one out of four tests of cognition. These tests were chosen to be representative of multiple cognitive domains as well as to be of slightly higher difficulty, and included CVLT-II Long-Delay Free Recall, WAIS-IV Digit Span Backward, D-KEFS Number-Letter Switching, and D-KEFS Color-Word Interference Inhibition.

Sampling procedures. Participants were recruited from a research participant registry maintained by UCCS. Randomly generated numbers were matched with participant ID numbers in the registry database to determine the order in which potential participants were contacted. They were contacted by phone, briefly informed of the nature of the study (including time commitment and financial compensation), and invited to participate or decline. Flyers were also posted around a local building serving primarily older adults; individuals who expressed interest were contacted by phone and
otherwise subject to the same recruitment procedures as those contacted from the participant registry. Potential participants were contacted until an adequate sample size was reached (based on a priori power analysis and previous studies of stereotype threat and diagnosis threat). Participants were randomly assigned to condition and the examiner was blind to condition (via experimental procedures detailed below). All testing was conducted individually with each participant in private research rooms on campus.

**Sample characteristics.** A total of 103 community-dwelling older adult participants were recruited. As noted above, initially, participants were to be excluded if they performed two standard deviations below a demographics-based mean score on more than one of four tests (a test of long-delay free recall, a test of working memory, a test of set-shifting, and a test of inhibition). However, this criterion yielded no exclusions (despite concern about cognition for some participants based on their test performance and self-report of recent change in cognition), and was deemed insufficiently stringent. Using more stringent criteria of 1.5 standard deviations for long-delay free recall and 1.3 standard deviations for all other tests (values were chosen based on available normative data), one participant was excluded for scores on two measures falling below these cutoffs. An additional participant was excluded for a score greater than two standard deviations below their demographics-based mean score on long-delay free recall (raising concerns about their ability to retain instructions).

The final sample therefore consisted of 101 older adults ranging in age from 60 to 93 ($M = 72.38, SD = 6.68$; one participant declined to respond). Of these, 74 participants were women (73.3%). The vast majority ($n = 97, 96.0\%$) were White/Non-Hispanic, with two participants (2.0%) identifying as Latino/Hispanic and two identifying as Other
(2.0%), but declining to specify further. Fifty-seven participants (56.4%) characterized their partner status as married/partnered, 19 as divorced (18.8%), 19 as widowed (18.8%), and 6 as single (5.9%). In terms of education level, one participant (1.0%) did not complete high school, 13 participants (12.9%) earned a high school diploma or GED, 26 participants (25.7%) had some college with no Bachelor’s degree, 16 participants (15.8%) held a Bachelor’s degree, 10 participants (9.9%) completed some post-graduate work with no degree, and 34 participants (33.6%) had a Master’s degree or higher; one participant declined to respond. The majority of participants identified as Christian denominations, with 21 (20.8%) identifying as Catholic and 43 (42.6%) identifying as Protestant. Other belief systems represented included Agnostic or Atheist \((n = 13, 12.9\%)\), Buddhist \((n = 3, 3.0\%)\), and Jewish – Nonorthodox \((n = 1, 1.0\%)\). Eighteen participants (17.8%) specified an “Other” religious/spiritual affiliation, and two (2.0%) declined to respond. Overall, participants considered themselves to be in relatively good health; on a scale from 1 (very poor health) to 7 (excellent health), only 6 (5.9%) rated themselves as less than somewhat healthy \((M = 5.49, SD = 1.18)\). A more detailed presentation of relevant demographic variables by group is presented in Table 1.

**Materials**

This study explored the impact of different threat manipulations on multiple cognitive outcomes assessed via commonly used clinical measures, detailed below. Self-reported variables including perceived changes in cognitive functioning, dementia worry, perceived stereotype threat, aging stereotype self-relevance and subjective age, perceived risk of dementia, and self-rated effort were also assessed. Additionally, a manipulation
Table 1

Demographic Characteristics by Condition

<table>
<thead>
<tr>
<th>Variables</th>
<th>Control</th>
<th>Stereotype Threat</th>
<th>Diagnosis Threat</th>
<th>Total Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>Range</td>
<td>M</td>
</tr>
<tr>
<td>Age</td>
<td>72.12</td>
<td>5.48</td>
<td>62–89</td>
<td>72.15</td>
</tr>
<tr>
<td>Subj. Age</td>
<td>59.22</td>
<td>13.60</td>
<td>16–80</td>
<td>60.06</td>
</tr>
<tr>
<td>Old Age Begins</td>
<td>78.81</td>
<td>9.21</td>
<td>0–95</td>
<td>81.67</td>
</tr>
<tr>
<td>Baseline DW</td>
<td>1.76</td>
<td>0.86</td>
<td>0–3</td>
<td>1.73</td>
</tr>
<tr>
<td>Self-Rated Health</td>
<td>5.56</td>
<td>1.02</td>
<td>3–7</td>
<td>5.18</td>
</tr>
<tr>
<td>Gender (% female)</td>
<td>67.65%</td>
<td>72.72%</td>
<td>79.41%</td>
<td>73.27%</td>
</tr>
<tr>
<td>AD Fam. Hx.</td>
<td>50.00%</td>
<td>57.57%</td>
<td>58.82%</td>
<td>55.45%</td>
</tr>
</tbody>
</table>

Note. Subj. Age = subjective age; Old Age Begins = Participant estimate of the age at which “old age” begins; Age Relevance = participant age minus their estimate of the age at which “old age” begins (negative number indicates participant feels younger than age at which old age begins); Baseline DW = baseline dementia worry, assessed prior to the study; AD Fam Hx = presence or absence of family history of dementia. The following variables had missing data: Age (one Control participant declined to answer); Subj. Age (two Control participants declined to answer); Old Age Begins (two Control participants declined to answer); Age Relevance (could not be calculated for three Control participants due to missing data).
check was included. Finally, participants completed a demographics form. These measures are described in detail below.

**Experimental manipulation.** Instructions were based on those used by Suhr and Gunstad (2002) to induce diagnosis threat, though some non-essential components of instructions were simplified to lower the reading level. Instructions varied between groups as follows.

*Neutral/control condition.* This condition made no reference to aging or disease processes. Participants read the following:

When the researcher returns to the room, s/he will ask you to complete a brief set of common neuropsychological tests. Some of the tests are easy, some are hard. Please try your best. Questions about tests will be answered following the testing.

*Age-based stereotype threat condition.* This condition emphasized the recruitment of participants based on their age and the expectation of cognitive decline with age, but made no reference to potential pathological causes of cognitive changes. Participants read the following:

It is widely assumed that older adults (age 60+) show cognitive changes compared to younger adults. Declines in areas such as attention, memory, and speed of information processing are commonly assumed—though other declines are sometimes presumed. This study examines the role that aging of the brain may play in these cognitive areas to better understand the nature of growing older.

When the researcher returns to the room, s/he will ask you to complete a brief set of common neuropsychological tests. These tests will assess skills such as attention, memory, speed of information processing, etc. Some of the tests are...
easy, some are hard. Please try your best. Questions about tests will be answered following the testing.

**Dementia-related diagnosis threat condition.** This condition emphasized the use of neuropsychological assessment instruments to assess for AD and other dementias, but did not explicitly discuss the participant’s age as relevant to the study. Participants read the following:

A growing number of studies find that many people with Alzheimer’s disease and other dementias show cognitive deficits on neuropsychological tests. Deficits in areas such as attention, memory, and speed of information processing are common in Alzheimer’s disease and other dementias—though other deficits sometimes emerge. This study examines how everyday people perform on tests typically used to assess for Alzheimer’s disease and other dementias. When the researcher returns to the room, s/he will ask you to complete a brief set of common neuropsychological tests. These tests will assess skills such as attention, memory, speed of information processing, etc. Some of the tests are easy, some are hard. Please try your best. Questions about tests will be answered following the testing.

All conditions’ instructions ended with the following statements: “For ethical reasons, the examiner will be unable to provide feedback on your individual performance. Please place these instructions back in the envelope and do not tell the examiner what you have read.”
Cognitive measures. Specific tests of cognitive functioning were chosen to assess multiple cognitive domains, and furthermore were selected based on their common use among practicing clinical neuropsychologists (as reported in Rabin et al., 2016).

Learning and memory. The California Verbal Learning Test Second Edition (CVLT-II; Delis, Kramer, Kaplan, & Ober, 2000) is an auditory list-learning test and delayed recall/recognition test. Test-takers are asked to immediately recall as many words as possible from a verbally presented 16-word list (List A) each trial over five learning trials. They are then given a single trial to recall as many words as possible from a separate 16-word list (List B). Next they complete a Short-Delay Free Recall trial in which they are asked to freely recall as many words as possible from List A, and subsequently administered a Short-Delay Cued Recall trial in which they are provided category cues (e.g., “Tell me all the words from the first list that are vegetables”) for each of the four semantic categories that compose the list. After an approximately 20-minute delay, participants complete a Long-Delay Free Recall and Long-Delay Cued Recall trial. Immediately after this, they complete a Long-Delay Yes/No Recognition trial in which they must indicate whether a given word was or was not on List A. The recognition trial includes 48 words comprising words from List A, words from List B, non-list words that fit List A semantic categories, and non-list words that do not fit any List A semantic categories. In this study, a recognition discriminability score was generated by subtracting number of false positive responses from number of correctly endorsed list words. This study’s hypotheses used Trials 1-5 Total Recall (higher scores indicate better learning and immediate recall), Long-Delay Free Recall (higher scores reflect better delayed recall), and Long-Delay Yes/No Recognition discriminability (higher scores
represent better ability to differentiate between target words and foils). In addition, participants were also asked to complete a Long-Delay Forced-Choice Recognition trial (administered 10 minutes after completing the initial recognition trial). In the forced-choice paradigm, participants are verbally presented 16 pairs of words, each of which contains one List A word and one semantically-unrelated non-list word. When each pair is given, the test-taker is asked to identify which word was from List A. In clinical use, this trial acts as a performance validity measure; in the context of this study, it was included as a proxy for participant effort (such that lower scores represent lower effort).

According to Delis et al. (2000), scores on the CVLT-II demonstrated good reliability as evidenced by high split-half reliability ($r = .94$ for the entire standardization sample; $rs = .92$ for adults aged 60-69, 70-79, and 80-89). Reliability was also found to be comparable for individuals at the low ($r = .87$), middle ($r = .88$), and high ($r = .89$) end of the performance distribution. Additionally, test-retest reliability for total recall was .82 over a 21-day period between testing. As would be expected, CVLT-II scores correlate moderately negatively with age ($r = -.51$). Scores on CVLT-II variables correlated well with their counterparts on the well-validated original CVLT, including Trials 1-5 Recall ($r = .76$), Long-Delay Free Recall ($r = .78$) and Total Recognition Discriminability ($r = .74$). Factor analysis (using principal components methods) of major CVLT-II variables among both a non-clinical and clinical sample suggested similar factor loadings across five to six factors including General Verbal Learning, Response Discrimination, Primacy-Recency Effects, Organizational Strategies, Recall Efficiency, and Total Learning Slope (non-clinical sample only). Each solution accounted for approximately 76% of the variance in their responsive samples (Delis et al., 2000). Subsequent research (Donders,
2008) criticized the methods used to identify these factors and conducted a confirmatory factor analysis to explore the validity of four hypothesized models across three age groups (young, 16-30 years; middle, 31-60 years; old, 61-89 years). Results supported a four-factor model including Attention Span, Learning Efficiency, Delayed Memory, and Inaccurate Memory, though the author cautions that this model was less robust among older adults. Nevertheless, the CVLT-II remains one of the most widely used memory assessment instruments in clinical neuropsychology (Rabin et al., 2016).

**Information processing speed, attention, and executive functioning.** The Delis-Kaplan Executive Function System (D-KEFS; Delis, Kaplan, & Kramer, 2001a, 2001b) Trail Making Test (TMT) includes five conditions that assess information processing speed, attention, and aspects of executive functioning (set-shifting, divided attention). In Condition 1, Visual Scanning, participants are presented with an array of circles, each filled with a number, and asked to find and mark through every 3. In Condition 2, Number Sequencing, 16 circles with numbers from 1 through 16 are scattered across the page and participants are asked to connect them in numerical sequence. Similarly, in Condition 3, Letter Sequencing, 16 circles with letters from A to P must be sequentially connected. In Condition 4, Number-Letter Switching, participants must switch between numbers and letters (e.g., connect 1 to A, A to 2, 2 to B, B to 3, and so on), starting at the number 1 and ending at the letter P. Condition 5, Motor Speed, requires rapidly tracing a dotted line connecting blank circles, beginning at a circle marked “Start” and ending at one marked “End.” Each condition includes a short practice trial (with feedback) to ensure participants understand the task. Raw scores are calculated as how many seconds were required to complete a given condition (such that higher raw scores represent poorer
performance). Conditions 1-3 and 5 are discontinued if not completed by 150 s; Condition 4 is discontinued if not completed by 240 s. Outcome measures of interest in this study included the mean raw score (i.e., seconds to complete) between the Number Sequencing and Letter Sequencing trials (as a measure of visuospatial processing speed and attention), and raw score on the Number-Letter Sequencing condition (as a measure of executive functioning, specifically set-shifting/divided attention).

According to standardization sample data reported by Delis et al. (2001b), internal consistency for sequencing portions of the test—calculated by treating Number Sequencing and Letter Sequencing conditions as split halves of the same test and correlating the two—was generally good in the age range the present study will evaluate ($r_{age \ 60-69} = .80$, $r_{age \ 70-79} = .60$, $r_{age \ 80-89} = .77$). In the 50-89 year age group, test-retest reliability coefficients were $r = .63$ for Visual Scanning, $r = .37$ for Number Sequencing, $r = .70$ for Letter Sequencing, $r = .55$ for Switching, and $r = .74$ for Motor Speed with a mean of 25 days between administrations. Evidence for validity of the D-KEFS tests in identifying executive deficits in numerous populations has been documented (see Delis, Kramer, Kaplan, & Holdnack, 2004 for populations and references). Additionally, performance on the D-KEFS TMT Switching condition has been shown to be a good predictor of everyday functioning (i.e., independence in basic activities of daily living such as eating and instrumental activities of daily living such as grocery shopping) among cognitively intact older adults (Mitchell & Miller, 2008). Correlations between the D-KEFS TMT and CVLT-II were evaluated by Delis et al. (2001b). As might be expected, the Visual Scanning condition did not correlate with CVLT-II total immediate recall, delayed free recall, or recognition discriminability, while the more cognitively
complex Switching condition did correlate (modestly) with total immediate recall ($r = .37$), long delay free recall ($r = .30$), and recognition discriminability ($r = .38$), with correlations between CVLT-II and the remaining three conditions falling in between the two. This pattern of correlations suggests both tests assess some aspects of general cognitive functioning, but without tremendous overlap in domains assessed.

**Attention and working memory.** The Wechsler Adult Intelligence Scale Fourth Edition (WAIS-IV; Wechsler, 2008) Digit Span subtest assesses attention and working memory. In Digit Span Forward, participants are verbally presented series of digits (increasing in length from sequences of two digits to sequences of nine) and asked to repeat the series verbatim. In Digit Span Backward, the presentation is similar but participants are instead asked to repeat the series in reverse order (e.g., 1-3 becomes 3-1). (An additional trial—Digit Span Sequencing, in which participants are asked to repeat the series in ascending numerical order—was not administered in the present study.) Each task is discontinued when a participant fails two items of the same sequence length. Higher scores indicate better performance (i.e., greater attentional and working memory capacity). Additionally, in this study a reliable digit span score was generated for each participant. This score added the longest digit span a participant was able to complete for two trials on Digit Span Forward with the longest span completed for two trials on Digit Span Backward (yielding a score between 0 and 17). Reliable digit span is often used as an embedded performance validity measure, where scores of either 6 or 7 or less are thought to represent poorer effort and/or potential malingering (Greiffenstein, Baker, & Gola, 1994; Schroeder, Twumasi-Ankrah, Baade, & Marshall, 2012).
Reliability and validity data reported in Wechsler (2008b) suggested scores on Digit Span have demonstrated excellent reliability across all age groups, with values of .92 for the 55-64-year-old age group, .93 for 65-69 years, .94 for 70-74 years, .93 for 75-79 years, and .92 for 80-84 and 85-90 years. Reliability coefficients for those with Mild Cognitive Impairment (.93) and mild AD (.95) were comparable as well, suggesting consistent reliability among both cognitively intact and impaired individuals. Digit Span also demonstrated good test-retest reliability in both the entire sample ($r = .82$) and among the age range of interest for this study ($r = .89$ for ages 55-69, $r = .84$ for ages 70-90). Among intercorrelations between all WAIS-IV subtests, evidence of convergent and discriminant validity for Digit Span was suggested by the subtest’s highest correlations with other tests of working memory such as a comparable letter-number sequencing task ($r = .69$) and a test of mental arithmetic ($r = .60$) and lower but still significant correlations with tests that measure different aspects of the same presumed underlying general intelligence factor such as a psychomotor coding task ($r = .45$) and a test of fund of knowledge ($r = .43$). Correlations between Digit Span subtasks ranged from .42 (between Digit Span Forward and Sequencing) and .53 (between Digit Span Forward and Backward). Additionally, scores on Digit Span have been shown to correlate modestly with the combined Number and Letter Sequencing scores on the D-KEFS TMT ($r = .38$) and more strongly with the Switching trial ($r = .53$), which requires greater attentional and working memory resources to maintain set. Scores on Digit Span additionally correlated positively but not excessively with scores on CVLT-II variables including Trials 1-5 recall ($r = .36$), long delay free recall ($r = .33$), and total recognition discriminability ($r = .31$).
Executive functioning (inhibition). The D-KEFS Color-Word Interference Test (CWIT; Delis et al., 2001a, 2001b) is a test of cognitive inhibition modeled after the original Stroop (1935) test. In Condition 1, Color Naming, test-takers are presented a page with red, blue, and green squares in several rows and columns. They are asked to name the colors as quickly as possible. In Condition 2, Word Reading, participants read aloud as quickly as possible from a page with the words red, blue, and green printed in black ink and arrayed again in rows and columns. In Condition 3, Inhibition, color words are printed in a different color ink (e.g., the word red printed in blue ink). Here test-takers are asked to name the color of the ink the word is printed in, rather than read the word aloud. (The D-KEFS CWIT additionally adds a Condition 4, Inhibition/Switching which follows Condition 3 save for the addition of color words in boxes interspersed in the word array; in these cases, participants are asked to read the word, rather than name the ink color. This trial was not administered.) For all conditions, the primary outcome measure was seconds taken to complete the trial (with a discontinue criterion of 90 s for Color Naming and Word Reading, and 180 s for Inhibition), such that higher scores represented poorer executive functioning. Errors (included both self-corrected and uncorrected errors) were also tallied in each condition.

According to Delis, Kaplan, and Kramer (2001b), internal consistency for the two baseline conditions—calculated by treating Color Naming and Word Reading conditions as split halves of the same test and correlating the two—was good in the age range this study evaluates ($r_{age\ 60-69} = .81$, $r_{age\ 70-79} = .81$, $r_{age\ 80-89} = .77$). In the 50-89 year age group, test-retest reliability coefficients were $r = .56$ for Color Naming, $r = .56$ for Word Reading, $r = .50$ for Inhibition, and $r = .57$ for Inhibition/Switching with a mean of 25
days between administrations. Additionally, practice effects were observed with improvement in scores across that time period. These values were lower than for other age groups and the sample as a whole, but still acceptable. Similar to data reported for the D-KEFS TMT, correlations with CVLT-II variables were higher in the more cognitively taxing Inhibition condition ($r = .35$ with CVLT-II immediate recall total, $r = .24$ with long delay free recall, and $r = .23$; Delis et al., 2001b).

Self-report Measures

*Dementia worry pre-measure.* To assess baseline dementia worry prior to study participation, an experimenter-created measure was used during recruitment. To avoid alerting participants to the nature of this study, a question related to anxiety about developing dementia was embedded within other health anxiety questions. For this measure, the examiner informed participants, “I would like to ask how concerned you are about personally developing a few different health issues at any point in the future.” Participants were then given the option to indicate that they were *very concerned,* *somewhat concerned,* *not very concerned,* or *not at all concerned* about (a) losing eyesight, (b) losing mobility or flexibility, (c) developing Alzheimer’s disease, (d) developing cardiovascular disease/heart disease, and (e) developing general problems with their emotional health, such as depression or anxiety. Each item was scored from 0 (*not at all concerned*) to 3 (*very concerned*) such that higher scores represent higher level of concern. The only question of interest for the purpose of this study was participant rating of their level of concern about the possibility of developing AD.

*Self-rated cognitive functioning.* In order to report perceived changes in cognitive functioning, participants completed the short form of the Everyday Cognition
(ECog) scale, the ECog-12 (Farias et al., 2011), based on the original 39-item ECog (Farias et al., 2008). Although the original ECog was developed and psychometrically validated for informant report, recent research has also made use of a self-report version of the scale (e.g., Amariglio et al., 2012; Edmonds, Delano-Wood, Galasko, Salmon, & Bondi, 2014). The original ECog included 39 items assessing change in functioning over the last 10 years in six cognitive domains: memory (sample item: “Remembering the current date or day of the week”), language (“Communicating thoughts in a conversation”), visuospatial abilities (“Reading a map and helping with directions when someone else is driving”), and three aspects of executive functioning including planning (“Anticipating weather changes and planning accordingly”), organization (“Keeping living and work space organized”), and divided attention (“Doing two things at once”). Responses are provided using a scale of 1 = better or no change, 2 = questionable/occasionally worse, 3 = consistently a little worse, 4 = consistently much worse. Using Item Response Theory analysis, items for the ECog-12 were chosen by selecting the items with the highest and lowest item difficulties from each cognitive domain in order to maximize discrimination of change in functioning at the widest possible range of levels of functioning. Responses to the original ECog may be evaluated as either a total mean score or domain-specific mean scores; responses to the ECog-12 are evaluated as a total mean score, with higher scores representative of greater perceived decline in cognitive functioning. The shortened form of the ECog was selected over the full 39-item version in order to minimize participant fatigue and maximize expected engagement in the task. Farias et al. (2011) suggest that the ECog-12 may be less useful
than the ECog in informant ratings of a dementia population owing to a floor effect in the ECog-12, but these measures are comparable for normal and MCI populations.

Data reported by Farias et al. (2011) suggest scores on the informant-report ECog-12 demonstrated excellent internal consistency ($\alpha = .96$). Additionally, evidence of convergent and discriminant validity was suggested by moderately strong correlations with other functional measures ($r_s = .64-.67$), weaker but still significant correlations with measures of cognitive functioning ($r_s = .44-.57$), and weakest correlations with demographic variables including age ($r = .17$) and education ($r = .10$).

**Dementia worry.** The Dementia Worry Scale (DWS; Kinzer & Suhr, 2015) includes 12 items that assess dementia worry. Participants rate items such as “My worries about dementia overwhelm me” and “When I can’t remember something, I find myself wondering whether I have dementia” using a scale from 1 (*not at all typical of me*) to 5 (*very typical of me*). Reported internal consistency for the DWS was excellent ($\alpha = .91$) and approximately 3-week test-retest reliability was .89. Additionally, scores on the DWS correlated moderately with scores on measures of depression ($r = .51$) and general worry ($r = .53$), suggesting convergent validity with related constructs but distinction from these constructs. Scores on the DWS also correlated moderately with belief in having a high likelihood of developing AD ($r = .61$) and modestly with memory concern ($r = .37$; Kinzer & Suhr, 2015). To facilitate interpretation, a mean item response score was calculated. Higher scores indicate greater levels of dementia worry.

**Perceived dementia risk.** Related to dementia worry, participants were asked to estimate their personal risk of developing dementia. In line with previous studies assessing perceived risk of AD (e.g., Health and Retirement Study, 2012; Linnenbringer,
Roberts, Hiraki, Cupples, & Green, 2010), participants were asked to provide percentage estimates of their perceived risk of developing dementia. Participants were asked, “Using a scale of 0-100 where 0 means no chance and 100 means absolutely certain, what are the chances that you will develop Alzheimer’s disease or another dementia sometime in the future?” To ascertain differences in personal perceived risk and perceptions of risk for peers, participants were also asked, “What are the chances (using a scale of 0-100 where 0 means no chance and 100 means absolutely certain) that the typical person your age will develop Alzheimer’s disease or another dementia sometime in the future?” Higher scores reflect greater perceived risk of developing AD.

**Self-rated effort.** To measure self-reported effort and test perception, participants were asked to complete a 5-item measure including the following questions taken from Suhr and Gunstad (2002): “How hard did you try on the tests?” (1 = *not hard at all* to 9 = *very hard*, higher scores represent greater reported effort), “How difficult did you find the tests?” (1 = *not difficult at all* to 9 = *very difficult*, higher scores indicate greater perceived difficulty), “How much pressure did you feel during testing?” (1 = *no pressure at all* to 9 = *very pressured*, higher scores reflect more perceived pressure), “How confident are you in your performance?” (1 = *not confident at all* to 9 = *very confident*, higher scores suggest greater confidence), and “How well did you do on the tests?” (1 = *very poorly* to 9 = *very well*, higher scores represent better perceived performance). In line with Suhr and Gunstad (2002), responses to each question will be considered individually (rather than analyzing results as a unitary scale), with responses to the first question used to answer Hypothesis 9.
**Perceived stereotype threat.** Although no gold standard for explicitly assessing stereotype threat exists, Steele and Aronson (1995) used a 5-item questionnaire that subsequent researchers have modified to fit different populations and outcome measures. For example, Chasteen et al. (2005) modified Steele and Aronson’s questions to assess age-related stereotype threat in the context of memory testing. To assess perceived stereotype threat in the present study, those questions were modified to better fit the broader cognitive domains assessed in this research. Participants used a scale from 1 = *strongly disagree* to 5 = *strongly agree* to respond to the following questions: “Some people feel I have less cognitive ability because of my age,” “Based on my age, people often underestimate my cognitive functioning,” “I often feel I have to prove to others that their perceptions of my cognitive functioning are wrong,” “The experimenter expected me to do poorly because of my age,” and “In experiments about cognitive functioning, people my age often face biased evaluations.” In order to ensure participants understand what is meant by *cognitive functioning* (and to best align with the domains identified in the manipulation instructions), these questions were introduced with a brief explanatory note: “The following questions ask for your opinion of others’ perception of your cognitive functioning, which includes things like attention, memory, and speed of information processing.” Mean item responses were calculated; higher scores represent greater perceived stereotype threat.

While not formally validated psychometrically, Chasteen et al. (2005) found consistently higher reports of stereotype threat among older adults than younger adults and responses on this scale demonstrated good internal consistency in two studies (Cronbach’s $\alpha = .74-.79$).
Demographics and age-related subjective variables. Finally, participants were asked to complete a basic demographics questionnaire including gender, ethnicity, partner status, education level, self-rated health (1 = very poor health to 4 = somewhat healthy to 7 = excellent health), religion, presence of personal or family history of AD or other dementias, and age.

In order to assess the relevance of aging stereotypes to individual participants’ identities (i.e., their level of identification as an older adult), participants were asked to answer the question “At what age do you believe ‘old age’ begins?” Similar measurements have previously been used to assess aging stereotype self-relevance (e.g., Kaufman & Elder, 2002; Levy, Zonderman, Slade, & Ferrucci, 2012; Molden & Maxfield, 2017). Additionally, given links between aging stereotypes and subjective age (e.g., Eibach, Mach, & Courtney, 2010), participants’ subjective age was measured. Consistent with Eibach et al., this was assessed via a question taken from the National Survey of Midlife in the United States: Participants read the statement, “Many people sometimes feel older or younger than they actually are,” and are then asked “What age do you feel at this moment?” They then complete the sentence “At this moment I feel ___ years old.” Each participant’s chronological age was subtracted from their subjective age, such that positive values represented feeling older than one’s chronological age and negative values represented feeling younger than one’s chronological age (with greater absolute values representing higher chronological/subjective age discrepancies).

Lastly, consistent with previous research on stereotype threat in which multiple choice questions were used as manipulation checks to assess participants’ understanding of instructions (e.g., Joanisse et al., 2013; Steele & Aronson, 1995), participants were
asked to answer the following multiple choice question (with responses based on manipulation scripts detailed below): “The purpose of this experiment was: (a) to complete a collection of both easy and hard neuropsychological tests, (b) to examine the role the aging of the brain plays in cognitive changes, or (c) to understand how everyday people perform on tests often used to diagnose Alzheimer’s disease and other dementias.” This manipulation check was administered prior to the measure of perceived stereotype threat in order to avoid participants’ over-endorsing answer choice B based on the age-emphasizing stereotype threat questions.

**Research Design and Procedure**

This study represented a between subjects experimental design in which threat induction was manipulated among experimental conditions. Main effects of experimental condition on both cognitive and self-report outcomes were evaluated.

**Measurement of baseline data.** During recruitment, participants were asked to respond to the 5-item pre-measure (detailed above) to assess baseline dementia worry. Participants were scheduled several days after recruitment to minimize suspicion regarding how the questionnaire related to the study activities.

**Introduction and experimental manipulation.** After obtaining informed consent, the examiner stepped out of the room to allow the participant to review test instructions contained in an envelope. Envelopes (one for each participant) were filled ahead of time (with one of three possible instructions included in each), shuffled by a third party, and assigned participant ID numbers afterward. This ensured that participants were randomly assigned to condition and that the experimenter was unaware of each participant ID number’s experimental condition until debriefing. Participants were
explicitly advised to return the instructions to the envelope before the examiner returned to the room. Participants were asked to slide the envelope under the door (or for less mobile participants, to open the door) when they finished reading the instructions in order to ensure their compliance with this request.

**Cognitive assessment.** After the examiner returned to the room, cognitive testing began (with examiner blind to participant condition). The CVLT-II Learning and Short-Delay Recall trials were administered first. In the 20-minute delay required before the Long-Delay Recall trials, participants completed WAIS-IV Digit Span and the D-KEFS TMT. If 19 min or more had elapsed, participants then completed the CVLT-II Long-Delay Recall trials and recognition testing; if 18 min or less had elapsed, participants instead completed D-KEFS CWIT trials. Those who did not complete this additional test prior to CVLT-II delayed tests were administered the task after completion of CVLT-II Long-Delay Yes/No Recognition. CVLT-II Forced-Choice Recognition was administered immediately after participants completed the self-report questionnaires described below.

**Self-report questionnaires.** Upon completion of cognitive testing, participants completed self-report questionnaires. These were administered electronically through Qualtrics, with paper versions (formatted identically to Qualtrics questionnaires) available for participants who were uncomfortable with computer use or in the event of technical difficulties ($n = 2$).

**Debriefing.** After participants filled out all questionnaires, they were thoroughly debriefed and provided the opportunity to ask questions about the study. Within the debriefing, participants were also typically asked to identify (or guess, if needed) which instructions they read based on descriptions of each instruction. Participants who
expressed an interest in feedback were reminded that given the nature of an experimental manipulation and that data must be analyzed anonymously, the testing was not diagnostic and the experimenter was unable to provide feedback on their performance. Interested participants were offered a referral list of local cognitive screening services and neuropsychologists.
CHAPTER III

RESULTS

Planned Statistical Analyses

To test the hypothesis that induction of stereotype threat and/or diagnosis threat would negatively impact cognitive functioning, a one-way multivariate analysis of variance (MANOVA) was conducted with condition (three levels: stereotype threat, diagnosis threat, and control) as the independent variable (IV) and 10 dependent variables (DVs): measures of memory functioning to answer Hypothesis 1, including CVLT-II Trials 1-5 Correct, CVLT-II Long-Delay Free Recall, and CVLT-II Recognition; measures of auditory attention and working memory to answer Hypothesis 2, including Digit Span forward and Digit Span backward; measures of visuospatial attention and processing speed to answer Hypothesis 3, including D-KEFS CWIT Color Naming, D-KEFS CWIT Word Reading, and the mean combination of D-KEFS TMT Number Sequencing and Letter Sequencing; and measures of executive functioning to answer Hypothesis 4, including D-KEFS TMT Switching and D-KEFS CWIT Inhibition. Based on recommendations suggested by Tabachnick and Fidell (2012), Wilks’ lambda was the planned statistical significance criterion, though Pillai’s trace was considered given its evidence of greater robustness to assumption violations. Planned follow-up analyses included univariate analyses of variance (ANOVAs) on DVs with the additional possibility of discriminant function analysis, as encouraged by Field (2009). All analyses were conducted on raw scores.
Testing the predicted mediational mechanisms of Hypothesis 5 depended on the nature of support for previous hypotheses. If data had supported predictions—that is, cognitive performance was negatively impacted to a similar magnitude in both stereotype threat and diagnosis threat conditions as compared to control—this would have required a moderated mediational analysis, whereby “the potency of the mediating process depends on the moderator” (Muller, Judd, & Yzerbyt, 2005, p. 856). In this case, perceived stereotype threat would have presented a potent mediator in the stereotype threat condition, but not the diagnosis threat condition. Similarly, perceived AD risk would have acted as a potent mediator for the diagnosis threat condition, but not the stereotype threat condition. However, should different magnitudes of effect on cognitive performance have been found between the two experimental conditions, this would have instead suggested a mediated moderation model. While conceptually these are slightly different, the statistical analyses required to assess them are the same (Muller et al., 2005). The method for assessing mediation outlined in Baron and Kenny (1986) and Muller et al. (2005) was planned.

To test hypotheses regarding the impact of experimental manipulation on self-report variables, four ANOVAs were conducted, one for each of four outcome measures: self-rated cognitive change as assessed by the ECog-12 (Hypothesis 6), dementia worry as measured by the DWS (Hypothesis 7), perceived stereotype threat as measured by Chasteen et al.’s (2005) questionnaire (Hypothesis 8), and self-rated effort as assessed via a question taken from Suhr and Gunstad (2002, 2005; Hypothesis 9). To reduce the risk of inflated familywise error rate, a Bonferroni-adjusted alpha of .0125 to accommodate four analyses was used. Although stronger differences were hypothesized between some
groups than others (e.g., between diagnosis threat and control conditions on self-reported cognitive change), given the lack of data comparing these different types of threats (and a complete lack of research on diagnosis threat in older adults), it was deemed important to explore all condition differences. Therefore, post hoc comparisons included examination of differences between all three experimental conditions.

**Participant Differences Between Groups**

One-way analysis of variance (ANOVA) indicated no significant difference in age between experimental conditions, $F(2, 97) = 0.13, p = .88$, partial $\eta^2 < .01$ (degrees of freedom for this analysis differ due to one participant declining to report age). Groups also did not differ as a function of self-rated health, $F(2, 98) = 1.78, p = .17$, partial $\eta^2 = .04$. There was no difference between gender representation in groups, $\chi^2 (2, N = 101) = 1.21, p = .55$, nor was there a difference between groups in whether or not participants had a family history of dementia, $\chi^2 (2, N = 101) = 0.62, p = .73$.

**Internal Consistency**

Scales were tested for internal consistency using Cronbach’s alpha. Scores on the ECog-12 demonstrated good reliability, $\alpha = .86$. Scores on the DWS evidenced excellent reliability, $\alpha = .93$. Though less consistent, scores on the measure of perceived stereotype threat nevertheless demonstrated acceptable reliability, $\alpha = .72$.

**Manipulation Check**

Results of a manipulation check consisting of a multiple choice question querying the participants’ understanding of the purpose of the study suggested a substantial portion of participants either did not accurately recall the instructions they read, or else suspected an alternative purpose for the study than that indicated in their instructions. Percentage of
participants endorsing each of the three possible study purposes is detailed in Table 2; note that one participant in the Diagnosis Threat condition did not answer this question. Results of a chi-square analysis suggested that some groups differed significantly from one another in their accuracy of multiple choice condition selection, $\chi^2 (2, N = 101) = 13.01, p = .001$. Follow-up chi-squares (using Yates Continuity Correction for $2 \times 2$ tables) were conducted to identify which groups were significantly different from one another in terms of participants’ accurate selection of condition. Results reflected that the control group differed significantly from the stereotype threat group, $\chi^2 (1, N = 67) = 11.01, p = .001$, and to a lesser extent from the diagnosis threat group, $\chi^2 (1, N = 68) = 3.79, p = .05$, such that those in the control group were less likely to accurately select their condition. The stereotype threat and diagnosis threat groups did not differ significantly from one another, $\chi^2 (1, N = 67) = 1.48, p = .22$.

Table 2

<table>
<thead>
<tr>
<th>Assigned Condition</th>
<th>Condition Endorsed by Participants in Manipulation Check</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
</tr>
<tr>
<td>Control</td>
<td>32.4% ($n = 11$)</td>
</tr>
<tr>
<td>Stereotype Threat</td>
<td>12.1% ($n = 4$)</td>
</tr>
<tr>
<td>Diagnosis Threat</td>
<td>8.8% ($n = 3$)</td>
</tr>
</tbody>
</table>

*Note. Accurate endorsement of condition is bolded for ease of visibility. One Diagnosis Threat participant who did not respond to this question is not represented in the table.*

In addition to a multiple choice question, during debriefing participants were also asked to identify their condition after having the different possible instructions described to them but before their condition was revealed. These results reflected more accurate recollection of instruction for the Control and Diagnosis Threat conditions as compared...
to accuracy rates estimated by the multiple choice manipulation check, but less accurate recall for the stereotype threat condition. Results of a chi-square analysis suggested groups were not statistically significant different from one another in terms of their accuracy of condition endorsement in debriefing, $\chi^2 (2, N = 101) = 4.93, p = .09$. These data are presented in Table 3.

Table 3

**Accuracy of Debriefing Manipulation Check by Condition**

<table>
<thead>
<tr>
<th>Assigned Condition</th>
<th>Control Endorsed by Participant in Debriefing</th>
<th>Stereotype Threat</th>
<th>Diagnosis Threat</th>
<th>Not Sure / Guessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>70.6% (n = 24)</td>
<td>0.0% (n = 0)</td>
<td>5.9% (n = 2)</td>
<td>23.5% (n = 8)</td>
</tr>
<tr>
<td>Stereotype Threat</td>
<td>27.3% (n = 9)</td>
<td>57.6% (n = 19)</td>
<td>6.1% (n = 2)</td>
<td>9.1% (n = 3)</td>
</tr>
<tr>
<td>Diagnosis Threat</td>
<td>2.9% (n = 1)</td>
<td>0.0% (n = 0)</td>
<td>82.4% (n = 28)</td>
<td>14.7% (n = 5)</td>
</tr>
</tbody>
</table>

*Note.* Accurate endorsement of condition is bolded for ease of visibility.

**Analyses – Hypotheses 1-4**

Hypotheses 1-4 predicted that cognitive outcome measures would be poorer among those in the stereotype threat and diagnosis threat conditions as compared to those in control condition. Each individual hypothesis considered a specific area of cognitive functioning, including memory functioning (Hypothesis 1), auditory attention and working memory (Hypothesis 2), visuospatial attention/processing speed (Hypothesis 3), and executive functioning (Hypothesis 4). Means, standard deviations, and score ranges for cognitive variables (by group and for the sample as a whole) are presented in Table 4.

A one-way multivariate analysis of variance (MANOVA) was performed with a three-level independent variable (condition: control, age-based stereotype threat, and dementia-related diagnosis threat) and ten dependent variables (DV$s$): CVLT Immediate
Table 4

Means, Standard Deviations, and Ranges of Cognitive Variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Control</th>
<th>Stereotype Threat</th>
<th>Diagnosis Threat</th>
<th>Total Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVLT Tot.</td>
<td>49.38</td>
<td>10.44</td>
<td>28–66</td>
<td>49.02</td>
</tr>
<tr>
<td>CVLT LDFR</td>
<td>11.38</td>
<td>2.89</td>
<td>4–16</td>
<td>4–16</td>
</tr>
<tr>
<td>CVLT Rec. DI</td>
<td>12.29</td>
<td>4.80</td>
<td>12.76</td>
<td>13.26</td>
</tr>
<tr>
<td>DSF</td>
<td>10.12</td>
<td>2.10</td>
<td>10.52</td>
<td>9.53</td>
</tr>
<tr>
<td>DSB</td>
<td>9.35</td>
<td>2.15</td>
<td>8.85</td>
<td>8.18</td>
</tr>
<tr>
<td>TMT Cond. 2&amp;3*</td>
<td>36.06</td>
<td>13.15</td>
<td>36.17</td>
<td>38.00</td>
</tr>
<tr>
<td>TMT Cond. 4*</td>
<td>81.44</td>
<td>28.01</td>
<td>93.76</td>
<td>89.94</td>
</tr>
<tr>
<td>CWIT Cond. 1*</td>
<td>29.76</td>
<td>6.23</td>
<td>32.64</td>
<td>33.24</td>
</tr>
<tr>
<td>CWIT Cond. 2*</td>
<td>22.09</td>
<td>4.39</td>
<td>22.70</td>
<td>23.62</td>
</tr>
<tr>
<td>CWIT Cond. 3*</td>
<td>63.38</td>
<td>15.60</td>
<td>62.21</td>
<td>63.71</td>
</tr>
<tr>
<td>CVLT FC</td>
<td>15.94</td>
<td>0.24</td>
<td>16.00</td>
<td>16.00</td>
</tr>
<tr>
<td>RDS</td>
<td>9.97</td>
<td>1.73</td>
<td>9.73</td>
<td>8.97</td>
</tr>
</tbody>
</table>

Note. CVLT Tot. = CVLT Trials 1-5 Total Recall; CVLT LDFR = CVLT Long-Delay Free Recall; CVLT Rec DI = CVLT Yes/No Recognition Discriminability (Hits minus False Positives); DSF = Digit Span Forward; DSB = Digit Span Backward; TMT Cond. 2\&3 = Mean of TMT Conditions 2 (Number Sequencing) and 3 (Letter Sequencing); TMT Cond. 4 = TMT Condition 4 (Number-Letter Switching); CWIT Cond. 1 = CWIT Condition 1 (Color Naming); CWIT Cond. 2 = CWIT Condition 2 (Word Reading); CWIT Cond. 3 = CWIT Condition 3 (Inhibition); CVLT FC = CVLT Forced Choice Recognition; RDS = Reliable Digit Span. *Higher values indicate poorer performance
Recall, CVLT Long-Delay Free Recall, CVLT Recognition Discriminability, WAIS-IV
Digit Span Forward, WAIS-IV Digit Span Backward, the mean of D-KEFS TMT
Conditions 2 (Number Sequencing) and 3 (Letter Sequencing), D-KEFS CWIT Condition
1 (Color Naming), D-KEFS CWIT Condition 2 (Word Reading), D-KEFS TMT
Condition 4 (Number-Letter Switching), and D-KEFS CWIT Condition 3 (Inhibition).

With the use of Wilks’ criterion, the combined DVs were significantly affected by
the experimental condition, Wilks’ lambda = .71, $F(20, 178) = 1.68, p = .04$. The results
suggested a modest impact of condition on the outcome variables, partial $\eta^2 = .16$.
Follow-up univariate ANOVAs revealed the significant impact of experimental condition
was on WAIS-IV Digit Span Backward, $F(2, 98) = 3.25, p = .04$, partial $\eta^2 = .06$, such
that those in the control condition ($M = 9.25, SD = 2.15$) outperformed those in both the
age-based stereotype threat ($M = 8.85, SD = 1.94$) and dementia-related diagnosis threat
($M = 8.18, SD = 1.60$) conditions. Pairwise comparisons using least significant
differences (LSD) methods suggested the source of this significant difference was
between the control and diagnosis threat conditions ($p = .01$); the stereotype threat
condition did not significantly differ from either the diagnosis threat ($p = .15$) or control
groups ($p = .28$).

An overall effect of experimental group on CWIT Condition 1 (Color Naming)
approached significance, $F(2, 98) = 2.83, p = .06$, partial $\eta^2 = .06$. Those in the control
condition ($M = 29.76, SD = 6.24$) completed this task more quickly than those in the
stereotype threat ($M = 32.64, SD = 6.08$) and diagnosis threat ($M = 33.24, SD = 6.91$)
conditions. Pairwise comparisons using LSD methods revealed a statistically significant
difference between the control condition and the diagnosis threat condition ($p = .03$),
while the difference between control and stereotype threat conditions approached significance ($p = .07$). The stereotype threat and diagnosis threat groups did not significantly differ from one another ($p = .70$).

There was no other evidence for statistically significant impact of experimental condition on any of the other dependent variables ($ps > .15$).

Unfortunately, data for several variables exhibited problematic skewness; additionally, univariate and multivariate outliers were present for several variables. Inspection of $z$-scores for all dependent variables revealed five participants whose scores were more than three standard deviations above or below the sample means on one or more measures. Two of these individuals were additionally characterized as multivariate outliers based on Mahalanobis distance that exceeded the critical value of 29.59 for a MANOVA with 10 dependent variables (a third univariate outlier was additionally quite close to the cutoff).

Elimination of the five outliers improved the most problematic violations of normality as evidenced by reductions in values of skewness divided by standard error of skewness, and data were subsequently reanalyzed using MANOVA. In this analysis, there was no omnibus effect of condition on cognitive variables, Wilks’ lambda = .74, $F(20, 168) = 1.40, p = .13$, partial $\eta^2 = .14$. An alternative MANOVA was conducted using the complete sample, but eliminating the three most problematic variables (CVLT Recognition Discriminability, which demonstrated strong negative skew; the mean of D-KEFS TMT Conditions 2 and 3, which demonstrated positive skew; and D-KEFS TMT Condition 4, which also demonstrated positive skew). Results of this analysis were significant, Wilks’ lambda = .76, $F(14, 184) = 1.89, p = .03$, partial $\eta^2 = .13$. Follow-up
analyses were comparable to those of the original MANOVA detailed above; namely, the control group performed significantly better on both WAIS-IV Digit Span Backward and CWIT Condition 1 (color naming) than the stereotype threat and diagnosis threat conditions.

In summary, results were not supportive of Hypotheses 1-4. Although there was evidence of a small omnibus effect of condition on cognitive tests, with minimal evidence of an experimental effect on a test of working memory (WAIS-IV Digit Span Backward) and a trend toward an effect on a single test of visuospatial attention/processing speed (CWIT Condition 1, color naming), these effects were quite small (partial $\eta^2 = .06$ for each individual subtest). In addition, results appeared to be influenced by the presence of outliers, the elimination of which also resulted in elimination of a statistically significant effect.

**Analysis – Hypothesis 5**

Hypothesis 5 predicted different mediational mechanisms driving stereotype threat versus diagnosis threat effects. Given a lack of substantial evidence of impact of the experimental manipulation on cognitive variables, no mediation analyses were conducted.

**Analysis – Hypothesis 6**

Hypothesis 6 predicted that those in the diagnosis threat condition would report the most changes in cognitive functioning over the past 10 years (as assessed via the ECog-12), that those in the control condition would report the fewest, and that those in the stereotype threat condition would report somewhere in between. Means, standard deviations, and score ranges for the ECog-12 as well as other subjective variables
discussed below are presented in Table 5. A one-way analysis of variance (ANOVA) was conducted to examine the impact of experimental condition on mean item response on the ECog-12. Although data were positively skewed, transformation of variables did not meaningfully change analysis of results; therefore, only results of analysis on untransformed variables are reported for greater ease of interpretation. Results suggested no significant difference in self-reported cognitive change between the experimental conditions, $F(2, 98) = 0.08, p = .92$, partial $\eta^2 < .01$. As such, Hypothesis 6 was not supported.

**Analysis – Hypothesis 7**

Hypothesis 7 predicted highest levels of dementia worry (as measured by the DWS) among those in the diagnosis threat condition and lowest levels in the control condition, with the stereotype threat group falling in between. A one-way ANOVA was conducted to examine the impact of experimental condition on mean item response on the DWS. Again, transforming variables to correct for positive skewness did not alter results, and therefore only analysis of untransformed variables is reported. Results were not indicative of any significant difference in levels of dementia worry between experimental groups, $F(2, 98) = 0.395, p = .68$, partial $\eta^2 = .01$. Hypothesis 7 was therefore not supported.

**Analysis – Hypothesis 8**

Hypothesis 8 predicted that those in the stereotype threat condition would report the highest levels of perceived stereotype threat, those in the control condition would report the lowest, and that those in the diagnosis threat group would fall in between the two. A one-way ANOVA was conducted on mean item response to a modified version of
Table 5

Means, Standard Deviations, and Ranges of Subjective Variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Control</th>
<th>Stereotype Threat</th>
<th>Diagnosis Threat</th>
<th>Total Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>Range</td>
<td>M</td>
</tr>
<tr>
<td>ECog</td>
<td>1.47</td>
<td>0.40</td>
<td>1.00–2.83</td>
<td>1.43</td>
</tr>
<tr>
<td>DWS</td>
<td>1.54</td>
<td>0.69</td>
<td>1.00–3.25</td>
<td>1.52</td>
</tr>
<tr>
<td>ST</td>
<td>1.76</td>
<td>0.61</td>
<td>1.00–3.60</td>
<td>2.26</td>
</tr>
<tr>
<td>Effort</td>
<td>7.50</td>
<td>2.22</td>
<td>1–9</td>
<td>7.97</td>
</tr>
<tr>
<td>Difficulty</td>
<td>5.35</td>
<td>1.94</td>
<td>1–9</td>
<td>5.64</td>
</tr>
<tr>
<td>Well</td>
<td>5.68</td>
<td>1.51</td>
<td>3–9</td>
<td>6.03</td>
</tr>
<tr>
<td>Self AD Risk</td>
<td>37.56</td>
<td>29.90</td>
<td>0–99</td>
<td>37.27</td>
</tr>
<tr>
<td>Other AD Risk</td>
<td>40.88</td>
<td>23.66</td>
<td>0–99</td>
<td>38.30</td>
</tr>
</tbody>
</table>

Note. ECog = Everyday Cognition Short Form mean item response; DWS = Dementia Worry Scale mean item response; ST = perceived stereotype threat measure mean item response; Effort = participants’ self-rated effort; Difficulty = participants’ rating of how difficult tasks were; Well = participants’ rating of how well they believe they did; Self AD Risk = participants’ rating of their perceived risk of developing dementia; Other AD Risk = participants’ rating of perceived risk of developing dementia of average person their age; Risk Diff. = Self AD Risk minus Other AD Risk
Chasteen et al.’s (2005) measure of stereotype threat to assess the effect of experimental condition on perceived stereotype threat. Only results of untransformed variables are reported as results were not changed by transformation for positive skewness. Results suggested a significant impact of experimental condition on perceived stereotype threat, $F(2, 98) = 4.81, p = .01$, partial $\eta^2 = .09$. Effect size suggested a small to moderate effect. Pairwise comparisons using LSD methods revealed that participants in the stereotype threat group ($M_{item} = 2.26, SD = 0.78$) reported significantly higher levels of perceived stereotype threat as compared to both the control group ($M_{item} = 1.76, SD = 0.61$), $p = .01$, and the diagnosis threat group ($M_{item} = 1.84, SD = 0.71$), $p = .02$. (These comparisons remained significant when using more stringent Bonferroni methods, which yielded $p = .01$ for stereotype threat versus control and $p = .05$ for stereotype threat versus diagnosis threat.) Using LSD methods, the diagnosis threat and control groups did not differ significantly from one another ($p = .66$). Hypothesis 8 was therefore supported.

**Analysis – Hypothesis 9**

Hypothesis 9 predicted highest levels of self-reported effort among those in the stereotype threat group and lowest levels among the diagnosis threat group, with those in the control condition falling in between. A one-way ANOVA was conducted to analyze the impact of experimental condition on self-reported effort as measured by the single item, “How hard did you try on the tests?” Although an alternative analysis was conducted on a version of this variable transformed to correct for highly negative skew, results did not differ from the analysis on the original untransformed variable and are therefore not reported. Although mean reported effort was highest among the stereotype threat group ($M = 7.97, SD = 1.68$) as compared to the diagnosis threat ($M = 7.53, SD =$
2.14) and control conditions \((M = 7.50, SD = 2.22)\) as predicted, results were not suggestive of a statistically significant difference between experimental groups on self-reported effort, \(F(2, 98) = 0.56, p = .57, \text{partial } \eta^2 = .01\).

As an alternative approach to assessing differences in effort among experimental groups, embedded validity measures typically used in clinical contexts to objectively assess effort were examined. The CVLT forced choice recognition measure demonstrated near zero variability, with only two participants achieving less than a perfect score on this measure; as such, this variable was not a sensitive measure of effort for this sample. Examination of participants’ reliable digit span, however, suggested a fairly normal distribution of scores and was analyzed further via one-way ANOVA. Omnibus results suggested a marginally significant effect of condition on reliable digit span, \(F(2, 98) = 3.07, p = .05, \text{partial } \eta^2 = .06\), representing a small effect. Pairwise comparisons using LSD methods revealed the diagnosis threat group \((M = 8.97, SD = 1.57)\) performed significantly worse than the control group \((M = 9.97, SD = 1.73)\), \(p = .02\). (Using more stringent Bonferroni pairwise comparison methods, \(p = .06\).) The diagnosis threat group also scored lower than the stereotype threat group \((M = 9.73, SD = 1.89)\), although this difference approached but did not achieve statistical significance \((p = .08)\). The stereotype threat group did not differ significantly from control \((p = .57)\).

In summary, analysis of participants’ self-reported effort did not support Hypothesis 9. However, examination of participants’ effort via embedded validity measures provided partial support for this hypothesis; effort as measured via reliable digit span was lowest among those in the diagnosis threat group as compared to the other two groups, but was not highest among those in the stereotype threat group.
Additional Analyses

Analyses using only those who accurately recalled condition. As detailed above, based on conversation with participants in debriefing, approximately 30% of the total sample did not accurately recall their condition after having the different possible instruction sets described to them. Analyses were therefore re-run using only participants who accurately recalled their condition. (Previously identified outliers on cognitive variables were also excluded from analysis.)

Results of a MANOVA assessing effect of experimental condition on cognitive variables remained nonsignificant (as compared to previous analyses excluding outliers), Wilks’ lambda = .70, $F(20, 112) = 1.07, p = .39$, partial $\eta^2 = .16$.

An ANOVA assessing impact of experimental group on self-reported cognitive change as measured by the ECog-12 remained nonsignificant, $F(2, 68) = 0.53, p = .59$, partial $\eta^2 = .02$.

An ANOVA examining effect of experimental group on dementia worry as measured by the DWS remained nonsignificant, $F(2, 68) = 0.08, p = .92$, partial $\eta^2 < .01$.

Results of an ANOVA exploring differences between groups on perceived stereotype threat were comparable to those conducted using the entire sample, $F(2, 68) = 5.29, p < .01$, partial $\eta^2 = .14$. As in the original analysis, those in the stereotype threat group ($M = 2.26, SD = 0.65$) reported significantly higher levels of perceived stereotype threat as compared to those in the diagnosis threat ($M = 1.74, SD = 0.66$) and control conditions ($M = 1.71, SD = 0.55$).

Results of a series of ANOVAs assessing the impact of experimental group on effort were also comparable to initial results using the entire sample. Once again, no
significant differences were found between groups on self-reported effort, $F(2, 68) = 0.30, p = .74$, partial $\eta^2 = .01$. However, as before a significant difference was found between groups on reliable digit span, $F(2, 68) = 4.12, p = .02$, partial $\eta^2 = .11$.

In summary, results did not appear to change meaningfully when analyses only included participants who accurately recalled the content of instructions in debriefing.

**Effects on global cognition.** A one-way ANOVA was conducted to test the effect of condition on an overall cognitive composite variable, calculated as the mean z-score of each of the 10 DVs (with TMT and CWIT scores recoded such that higher scores represented better performance). Results did not reflect a significant impact of experimental group on the cognitive composite measure, $F(2, 98) = 0.84, p = .44$, partial $\eta^2 = .02$.

**Multinomial logistic regression.** An alternative statistical analysis was conducted as a secondary way of assessing any impact of experimental condition on cognitive variables. Multinomial logistic regression was used to predict participants’ experimental condition based on their performance on cognitive tests. Data were not found to significantly violate assumptions of linearity in the logit. Examination of data for problems with multicollinearity reflected dependency observed between CVLT Total Recall and CVLT Long Delay Free Recall (as might be expected between immediate and delayed recall measures of the same word list) as well as between CWIT Conditions 1 (Color Naming) and 2 (Word Reading). Inspection of tolerance and variance inflation factor (VIF) statistics suggested multicollinearity was not problematic enough to be concerned about model bias, with the highest VIF value of 2.85 (where values of 10 and
above are cause for concern; Field, 2009) and lowest tolerance value of .35 (where values below .20 are cause for concern; Field, 2009).

In initial analyses, Pearson goodness of fit statistics suggested no significant discrepancies between the model’s predicted and observed values, \(X^2 (180, N = 101) = 192.67, p = .25\). Log-likelihood ratios (-2LL) decreased significantly from the baseline intercept-only model (-2LL = 221.90) to one including all predictors (-2LL = 188.01), suggesting significant improvement in explained variance, \(X^2 (20, N = 101) = 33.89, p = .03\), Nagelkerke \(R^2 = .32\). Variables that contributed significantly to the overall model were Digit Span Backward, \(X^2 (2, N = 101) = 7.23, p = .03\), and CWIT Condition 1 (Color Naming), \(X^2 (2, N = 101) = 7.84, p = .02\). Examination of parameter estimates (see Table 6) revealed that these same variables significantly differentiated between the diagnosis threat and control conditions; among the stereotype threat group, CWIT Condition 1 and CVLT Long-Delay Free Recall significantly differentiated from the control group, although the latter variable did not contribute in a statistically significant way to the overall model. Taken together, this model accurately predicted group membership for 55.9% of those in the control condition, 57.6% of those in the stereotype threat group, and 52.9% of those in the diagnosis threat group.

Examination of the data while excluding five previously identified outliers yielded a slight increase in explained variance as evidenced by a lower log-likelihood ratio (-2LL = 181.54), but eliminated any significant difference from a baseline intercept-only model (-2LL = 210.87) in terms of model fit, \(X^2 (20, N = 96) = 29.33, p = .08\), Nagelkerke \(R^2 = .30\). Variables’ overall contribution to the (nonsignificant) model remained basically unchanged; parameter estimates are presented in Table 7. When
analyses were conducted with the additional exclusion of those who did not accurately recall their condition in debriefing, an additional increase in variance explained was seen via a lower log-likelihood ratio (-2LL = 122.68), but again, the model remained nonsignificantly different from a baseline model, $X^2 (2, N = 68) = 24.51, p = .22$, Nagelkerke $R^2 = .34$.

Table 6

**Logistic Regression Results Using All Participants (N = 101)**

<table>
<thead>
<tr>
<th>Variable</th>
<th>B (SE)</th>
<th>Lower Bound</th>
<th>Odds Ratio</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stereotype Threat vs. Control</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-2.62 (3.62)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>CVLT Total</td>
<td>0.05 (0.05)</td>
<td>0.95</td>
<td>1.05</td>
<td>1.15</td>
</tr>
<tr>
<td>CVLT LDFR</td>
<td>-0.31 (0.14)*</td>
<td>0.55</td>
<td>0.73</td>
<td>0.97</td>
</tr>
<tr>
<td>CVLT Rec DI</td>
<td>0.12 (0.11)</td>
<td>0.91</td>
<td>1.13</td>
<td>1.39</td>
</tr>
<tr>
<td>DSF</td>
<td>0.20 (0.15)</td>
<td>0.90</td>
<td>1.22</td>
<td>1.65</td>
</tr>
<tr>
<td>DSB</td>
<td>-0.10 (0.17)</td>
<td>0.66</td>
<td>0.91</td>
<td>1.26</td>
</tr>
<tr>
<td>TMT Cond. 2&amp;3</td>
<td>-0.03 (0.03)</td>
<td>0.92</td>
<td>0.97</td>
<td>1.03</td>
</tr>
<tr>
<td>CWIT Cond. 1</td>
<td>0.18 (0.08)*</td>
<td>1.03</td>
<td>1.20</td>
<td>1.40</td>
</tr>
<tr>
<td>CWIT Cond. 2</td>
<td>-0.12 (0.11)</td>
<td>0.71</td>
<td>0.89</td>
<td>1.11</td>
</tr>
<tr>
<td>TMT Cond. 4</td>
<td>0.02 (0.01)</td>
<td>1.00</td>
<td>1.02</td>
<td>1.05</td>
</tr>
<tr>
<td>CWIT Cond. 3</td>
<td>-0.05 (0.03)</td>
<td>0.91</td>
<td>0.96</td>
<td>1.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>B (SE)</th>
<th>Lower Bound</th>
<th>Odds Ratio</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diagnosis Threat vs. Control</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.71 (3.52)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>CVLT Total</td>
<td>0.05 (0.05)</td>
<td>0.95</td>
<td>1.05</td>
<td>1.15</td>
</tr>
<tr>
<td>CVLT LDFR</td>
<td>-0.11 (0.14)</td>
<td>0.68</td>
<td>0.90</td>
<td>1.18</td>
</tr>
<tr>
<td>CVLT Rec DI</td>
<td>0.12 (0.11)</td>
<td>0.91</td>
<td>1.12</td>
<td>1.38</td>
</tr>
<tr>
<td>DSF</td>
<td>-0.07 (0.16)</td>
<td>0.68</td>
<td>0.93</td>
<td>1.27</td>
</tr>
<tr>
<td>DSB</td>
<td>-0.41 (0.17)*</td>
<td>0.48</td>
<td>0.67</td>
<td>0.92</td>
</tr>
<tr>
<td>TMT Cond. 2&amp;3</td>
<td>0.00 (0.03)</td>
<td>0.94</td>
<td>1.00</td>
<td>1.06</td>
</tr>
<tr>
<td>CWIT Cond. 1</td>
<td>0.18 (0.08)*</td>
<td>1.03</td>
<td>1.19</td>
<td>1.38</td>
</tr>
<tr>
<td>CWIT Cond. 2</td>
<td>-0.09 (0.11)</td>
<td>0.74</td>
<td>0.91</td>
<td>1.13</td>
</tr>
<tr>
<td>TMT Cond. 4</td>
<td>0.00 (0.01)</td>
<td>0.98</td>
<td>1.00</td>
<td>1.03</td>
</tr>
<tr>
<td>CWIT Cond. 3</td>
<td>-0.04 (0.03)</td>
<td>0.92</td>
<td>0.96</td>
<td>1.01</td>
</tr>
</tbody>
</table>

Note. CVLT Total = CVLT Trials 1-5 Total Recall; CVLT LDFR = CVLT Long-Delay Free Recall; CVLT Rec DI = CVLT Yes/No Recognition Discriminability, hits minus false positives; DSF = Digit Span Forward; DSB = Digit Span Backward; TMT Cond. 2&3 = mean score of TMT Conditions 2 (Number Sequencing) and 3 (Letter Sequencing); CWIT Cond. 1 = CWIT Condition 1 (Color Naming); CWIT Cond. 2 = CWIT Condition 2 (Word Reading); TMT Cond. 4 = TMT Condition 4 (Number-Letter Switching); CWIT Cond. 3 = CWIT Condition 3 (Inhibition).

* $p < .05$
Table 7

*Logistic Regression Results Excluding Outliers (n = 96)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>B (SE)</th>
<th>Lower Bound</th>
<th>Odds Ratio</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stereotype Threat vs. Control</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-2.88 (3.72)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>CVLT Total</td>
<td>0.06 (0.05)</td>
<td>0.96</td>
<td>1.06</td>
<td>1.17</td>
</tr>
<tr>
<td>CVLT LDFR</td>
<td>-0.32 (0.15)*</td>
<td>0.55</td>
<td>0.73</td>
<td>0.97</td>
</tr>
<tr>
<td>CVLT Rec DI</td>
<td>0.10 (0.13)</td>
<td>0.87</td>
<td>1.11</td>
<td>1.42</td>
</tr>
<tr>
<td>DSF</td>
<td>0.19 (0.16)</td>
<td>0.89</td>
<td>1.21</td>
<td>1.64</td>
</tr>
<tr>
<td>DSB</td>
<td>-0.14 (0.17)</td>
<td>0.63</td>
<td>0.87</td>
<td>1.22</td>
</tr>
<tr>
<td>TMT Cond. 2&amp;3</td>
<td>-0.01 (0.04)</td>
<td>0.92</td>
<td>0.99</td>
<td>1.06</td>
</tr>
<tr>
<td>CWIT Cond. 1</td>
<td>0.20 (0.08)*</td>
<td>1.04</td>
<td>1.22</td>
<td>1.44</td>
</tr>
<tr>
<td>CWIT Cond. 2</td>
<td>-0.14 (0.12)</td>
<td>0.69</td>
<td>0.87</td>
<td>1.10</td>
</tr>
<tr>
<td>TMT Cond. 4</td>
<td>0.01 (0.01)</td>
<td>0.99</td>
<td>1.01</td>
<td>1.04</td>
</tr>
<tr>
<td>CWIT Cond. 3</td>
<td>-0.04 (0.03)</td>
<td>0.91</td>
<td>0.96</td>
<td>1.02</td>
</tr>
<tr>
<td><strong>Diagnosis Threat vs. Control</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.25 (3.63)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>CVLT Total</td>
<td>0.06 (0.05)</td>
<td>0.96</td>
<td>1.06</td>
<td>1.17</td>
</tr>
<tr>
<td>CVLT LDFR</td>
<td>-0.13 (0.14)</td>
<td>0.67</td>
<td>0.88</td>
<td>1.17</td>
</tr>
<tr>
<td>CVLT Rec DI</td>
<td>0.09 (0.12)</td>
<td>0.87</td>
<td>1.10</td>
<td>1.38</td>
</tr>
<tr>
<td>DSF</td>
<td>-0.05 (0.16)</td>
<td>0.69</td>
<td>0.95</td>
<td>1.30</td>
</tr>
<tr>
<td>DSB</td>
<td>-0.45 (0.17)*</td>
<td>0.45</td>
<td>0.64</td>
<td>0.90</td>
</tr>
<tr>
<td>TMT Cond. 2&amp;3</td>
<td>0.01 (0.04)</td>
<td>0.94</td>
<td>1.01</td>
<td>1.08</td>
</tr>
<tr>
<td>CWIT Cond. 1</td>
<td>0.18 (0.08)*</td>
<td>1.02</td>
<td>1.20</td>
<td>1.41</td>
</tr>
<tr>
<td>CWIT Cond. 2</td>
<td>-0.12 (0.11)</td>
<td>0.71</td>
<td>0.89</td>
<td>1.11</td>
</tr>
<tr>
<td>TMT Cond. 4</td>
<td>0.00 (0.01)</td>
<td>0.97</td>
<td>1.00</td>
<td>1.03</td>
</tr>
<tr>
<td>CWIT Cond. 3</td>
<td>-0.02 (0.03)</td>
<td>0.93</td>
<td>0.98</td>
<td>1.03</td>
</tr>
</tbody>
</table>

*Note.* CVLT Total = CVLT Trials 1-5 Total Recall; CVLT LDFR = CVLT Long-Delay Free Recall; CVLT Rec DI = CVLT Yes/No Recognition Discriminability, hits minus false positives; DSF = Digit Span Forward; DSB = Digit Span Backward; TMT Cond. 2&3 = mean score of TMT Conditions 2 (Number Sequencing) and 3 (Letter Sequencing); CWIT Cond. 1 = CWIT Condition 1 (Color Naming); CWIT Cond. 2 = CWIT Condition 2 (Word Reading); TMT Cond. 4 = TMT Condition 4 (Number-Letter Switching); CWIT Cond. 3 = CWIT Condition 3 (Inhibition).

* *p < .05*

In sum, alternative analysis of the data using logistic regression largely mirrored results using MANOVA. Although initial results suggested statistical significance primarily driven by scores on Digit Span Backward and CWIT Condition 1, significance
was lost once outliers were accounted for, and results were not improved via inclusion of only those participants who accurately recalled their condition in debriefing.

**Inclusion of moderators.** Based on stereotype threat literature suggesting potential moderating effects of some variables, additional analyses were conducted to explore the potential impact of moderators. All moderator analyses were conducted as 2 (moderator of interest) by 3 (experimental group) MANOVAs for cognitive variable DVs, or 2 by 3 ANOVAs for subjective variable DVs. Although interactions were the primary effect of interest in these analyses, main effects are also described below when significant.

**Chronological age.** Given mixed evidence for potential different impact of threat effects depending on age, age was evaluated as a potential moderating variable. A median split was used to categorize participants as younger (71 years of age or less, \( n = 49 \)) or older (72 years of age or greater, \( n = 51 \)). Results of a MANOVA (excluding five participants deemed outliers on cognitive variables) did not reveal a statistically significant interaction between condition and age, Wilks’ lambda = .76, \( F(20, 160) = 1.20, p = .26 \), partial \( \eta^2 = .13 \).

A series of ANOVAs was conducted to assess whether age interacted with condition to affect any of the subjective variables. No significant interaction effects were found for any of these variables, including self-reported change in cognitive functioning as measured by the ECog-12, \( F(2, 94) = 2.19, p = .12 \), partial \( \eta^2 = .05 \); dementia worry, \( F(2, 94) = 1.51, p = .23 \), partial \( \eta^2 = .03 \); perceived stereotype threat, \( F(2, 94) = 0.97, p = .38 \), partial \( \eta^2 = .02 \); and self-rated effort, \( F(2, 94) = 0.30, p = 0.73 \), partial \( \eta^2 = .01 \).
Baseline dementia worry. Although dementia worry research is primarily cross-sectional and correlational, thus offering little evidence for hypotheses regarding how baseline dementia worry might impact cognitive functioning and the subjective variables considered in this study, it stands to reason that higher levels of dementia worry could potentially lead to greater susceptibility to threat effects. Participant responses to a pre-measure administered over the phone during scheduling were used to estimate baseline levels of dementia worry. In response to the question, “How concerned are you about developing Alzheimer’s disease?,” 10 participants (9.9%) rated themselves as not at all concerned, 32 (31.7%) rated themselves as not very concerned, 32 (31.7%) rated themselves somewhat concerned, and 27 (26.7%) rated themselves very concerned. To avoid untenably small cell sizes, participants were divided into two groups such that those who rated themselves as not at all or not very concerned were classified as “no or low baseline dementia worry” and those who rated themselves as somewhat or very concerned were classified as “mid or high baseline dementia worry.”

A MANOVA was conducted to explore any potential interaction effect between baseline dementia worry and experimental group. Results did not reflect a significant interaction, Wilks’ lambda = .86, F(20, 162) = 0.65, p = .87, partial η² = .08. Interestingly, results did suggest a modest main effect of baseline dementia worry on the DVs, Wilks’ lambda = .81, F(10, 81) = 1.94, p = .05, partial η² = .19; however, no follow-up univariate ANOVAs on each DV were statistically significant (ps > .07).

A series of ANOVAs was conducted to assess whether baseline dementia worry interacted with condition to affect any of the subjective variables. No significant interaction effects were found for any of these variables, including self-reported change
in cognitive functioning as measured by the ECog-12, $F(2, 95) = 0.76, p = .47$, partial $\eta^2 = .02$; dementia worry as measured by the DWS, $F(2, 95) = 0.87, p = .42$, partial $\eta^2 = .02$; perceived stereotype threat, $F(2, 95) = 0.38, p = .68$, partial $\eta^2 = .01$; and self-rated effort, $F(2, 95) = 0.35, p = .71$, partial $\eta^2 = .01$. Unsurprisingly, a significant main effect of baseline dementia worry was found on the DWS, $F(1, 95) = 29.23, p < .001$, partial $\eta^2 = .24$, such that those classified as mid or high baseline dementia worry had higher mean item responses on the DWS ($M = 1.74, SD = 0.70$) than those classified as low or no baseline dementia worry ($M = 1.14, SD = 0.16$).

**Family history of dementia.** Being aware of a family history of AD and other dementias can be assumed to potentially heighten one’s sensitivity to dementia-related thoughts and behaviors in oneself. As such, analyses examining effect of experimental group on cognitive variables and subjective variables of interest were conducted with the inclusion of family history of dementia as a possible moderator. However, there was no evidence of a significant interaction between family history and condition on neuropsychological test performance, Wilks’ lambda = .87, $F(20, 162) = 0.59, p = .92$, partial $\eta^2 = .07$.

With regard to subjective variables, there was no significant interaction effect of condition and dementia family history on self-reported change in cognitive functioning as measured by the ECog-12, $F(2, 95) = 0.13, p = .88$, partial $\eta^2 < .01$. There was no significant interaction effect of condition and family history on perceived stereotype threat, $F(2, 95) = 0.16, p = .85$, partial $\eta^2 < .01$. There was also no significant interaction between condition and dementia family history on self-reported effort, $F(2, 95) = 1.34, p = .27$, partial $\eta^2 = .03$. Although a main effect of family history on self-reported effort
approached significance, $F(1, 95) = 2.98, p = .09$, partial $\eta^2 = .03$, such that those with a family history rated their effort slightly higher ($M = 7.98, SD = 1.50$) than those without ($M = 7.27, SD = 2.49$), the effect size suggested this was not a particularly meaningful effect.

Although there was no significant interaction between family history of dementia and experimental group on dementia worry as measured by the DWS, $F(2, 95) = 0.44, p = .65$, partial $\eta^2 = .01$, there was a main effect of family history on dementia worry, $F(1, 95) = 5.39, p = .02$, partial $\eta^2 = .05$, representing a small effect. Those with a family history had a modestly higher mean item response on the DWS ($M = 1.61, SD = 0.64$) than those with no reported family history ($M = 1.34, SD = 0.55$).

**Education.** Participants’ reported education was converted to a dichotomized variable such that those who reported no more than a few years of college with no degree were classified as “low” education, and those with a Bachelor’s degree or higher were classified as “high” education. (Associate’s degrees were considered two years of education and included in the “low” education group.) MANOVA results did not reveal a statistically significant interaction between education and experimental condition on cognitive DVs, Wilks’ lambda = .77, $F(20, 170) = 1.22, p = .25$, partial $\eta^2 = .13$.

A series of ANOVAs were also conducted with subjective variables as DVs. There was no significant interaction between education and condition on mean item response on the ECog-12, $F(2, 94) = 0.19, p = .83$, partial $\eta^2 < .01$. There was also no significant interaction between education and experimental group on dementia worry as measured by the DWS, $F(2, 94) = 0.20, p = .82$, partial $\eta^2 < .01$, although there was a significant main effect of education on dementia worry, $F(1, 94) = 5.23, p = .02$, partial
$\eta^2 = .05$. Those with lower education endorsed higher levels of dementia worry ($M_{\text{item}} = 1.66$, $SD = 0.71$) than those with higher education ($M_{\text{item}} = 1.37$, $SD = 0.52$). There was no significant interaction between education and experimental group on perceived stereotype threat, $F(2, 94) = 0.15$, $p = .86$, partial $\eta^2 < .01$. Finally, there was also no interaction between education and condition on self-rated effort, $F(2, 94) = 0.54$, $p = .59$, partial $\eta^2 = .01$, although a main effect of education was observed, $F(1, 94) = 22.95$, $p = .02$, partial $\eta^2 = .06$. Individuals in the higher education group generally rated themselves as having tried harder ($M = 8.02$, $SD = 1.61$) than those in the lower education group ($M = 7.10$, $SD = 2.45$).

**Consideration of potential covariates.** Data were also re-analyzed with the inclusion of covariates that could potentially have impacted results.

**Relevance of age-related stereotypes.** Given suggestions that stereotype-related experimental effects may be dependent on relevance of stereotypes to a person’s identity (Levy, 2009; Molden & Maxfield, 2017), it was suspected that controlling for relevance of age-related stereotypes might reveal more significant impact of stereotype threat (and possibly diagnosis threat) effects on cognitive functioning and subjective variables. Responses to the question “At what age do you believe old age begins?” were subtracted from participants’ ages to yield a quantification of the relevance of age-related stereotypes to each participant. Positive numbers would suggest the participant considers themselves as past the point where “old age” begins; negative numbers suggest the participant does not consider themselves to have reached that point. (This value could not be generated for two participants who refused to answer this question/entered zero, or for a third participant who did not indicate her chronological age.) Mean responses suggested
the average participant considered him- or herself several years away from old age ($M = -7.41$, $SD = 13.12$), but also varied widely, ranging from -54.00 to 22.00.

A one-way ANOVA did not suggest a significant impact of experimental condition on the age relevance variable, $F(2, 95) = 0.89, p = .41$, partial $\eta^2 = .02$, indicating acceptability of this variable as a covariate. A multivariate analysis of covariance (MANCOVA) was therefore conducted assessing the impact of experimental condition on the 10 DVs when controlling for relevance of age stereotypes (excluding previously identified outliers). However, results were not meaningfully different from previous analysis without using a covariate, Wilks’ lambda = .71, $F(20, 162) = 1.49, p = .09$, partial $\eta^2 = .16$.

Hypotheses concerning subjective variables were also re-analyzed controlling for relevance of aging stereotypes. Results were not significantly different from previous analyses that did not make use of covariates. There was no significant impact of experimental group on self-reported change in cognition as measured by the ECog-12, $F(2, 94) = 0.13, p = .88$, partial $\eta^2 < .01$, nor was there an effect on dementia worry, $F(2, 94) = 0.27, p = .76$, partial $\eta^2 = .01$, nor was there an effect on participants’ self-reported effort, $F(2, 94) = 0.59, p = .56$, partial $\eta^2 = .01$. As before, there was a significant effect of experimental group on perceived stereotype threat, $F(2, 94) = 5.56, p = .01$, partial $\eta^2 = .11$, such that those in the stereotype threat group endorsed higher levels of stereotype threat as compared to those in the control condition, $p = .002$, and those in the diagnosis threat condition, $p = .02$ (pairwise comparisons are not adjusted for multiple comparisons).
In summary, controlling for the relevance of age-related stereotypes based on participants’ perception of their own closeness/distance to old age did not meaningfully impact original hypothesis testing.

**Task difficulty.** Based on evidence suggesting threat is more likely to be operative on difficult tasks as compared to easier tasks (Spencer et al., 1999), an additional MANCOVA was conducted while controlling for participant’s ratings of how difficult they found the tasks. There was no evidence that task difficulty ratings were affected by experimental group, \( F(2, 98) = 0.72, p = .49, \text{partial } \eta^2 = .01 \). MANCOVA results (excluding outliers) did not indicate that controlling for participants’ perception of task difficulty meaningfully changed effect of condition on DVs, Wilks’ lambda = .74, \( F(20, 166) = 1.35, p = .16, \text{partial } \eta^2 = .14 \).

**Exploratory analyses.** Some variables were included in the study owing to a lack of abundant research on their susceptibility to threat effects. No formal hypotheses were made for these variables, but they were analyzed as potential areas for future research.

**Subjective age.** As an exploratory analysis, data were analyzed to identify whether subjective age (as measured by numerical response to the question, “Many people sometimes feel younger or older than they actually are. What age do you feel at this moment?”) was impacted by experimental condition. Two participants declined to answer this question, and one participant who completed questionnaires on paper rather than on computer (where numerical responses were forced) responded with “teenager”; her response was entered as 16. Responses on this measure ranged from 16 to 85 (\( M = 59.76, SD = 11.80 \)), but did not differ between experimental groups, \( F(2, 96) = 0.05, p = .95, \text{partial } \eta^2 < .01 \). Condition additionally did not influence the difference between
chronological and subjective age, \(F(2, 95) = 0.07, p = .94\), partial \(\eta^2 < .01\). On average, participants reported a subjective age 12.69 years younger (\(SD = 11.87\)) than their chronological age.

**Perceived risk of AD.** An additional exploratory analysis entailed consideration of whether perceived risk of AD might have been impacted by experimental condition. Participants’ estimates of their own risk of developing AD using a scale from 0 (*no chance*) to 100 (*absolutely certain*) ranged from the full spectrum of 0 to 100. Despite the broad range, participants generally endorsed relatively low risk for themselves (\(M = 33.56, SD = 28.87\)) and only slightly higher risk for others (\(M = 38.22, SD = 21.33\)), although this difference was statistically significant, \(t(100) = -2.21, p = .03\), two-tailed. On average, participants perceived their own risk as 4.65 percentage points lower than the average person their age (\(SD = 21.13\)). This difference was primarily driven by participants with no family history of dementia, who rated their perceived risk for themselves (\(M = 20.93, SD = 26.59\)) much lower than their perceived risk for the average person their age (\(M = 33.60, SD = 21.24\)). Those with a positive family history rated risks as generally equivalent (\(M = 43.71, SD = 26.72\) for self; \(M = 41.93, SD = 20.84\) for others). In a one-way ANOVA, family history significantly affected both perceived risk for self, \(F(1, 99) = 18.22, p < .001\), partial \(\eta^2 = .16\), and the difference between perceived risk for self and perceived risk for other, \(F(1, 99) = 13.08, p < .001\), partial \(\eta^2 = .12\).

Results of one-way ANOVA suggested participants perceived AD risk for themselves was not impacted by experimental manipulation, \(F(2, 98) = 1.80, p = .17\), partial \(\eta^2 = .04\). Experimental group additionally did not impact the difference between
participants’ perceived AD risk for themselves and their perceived risk for the average person their age, \( F(2, 98) = 1.46, p = .24\), partial \( \eta^2 = .03\).

**Correlations.** Cognitive, subjective, and demographics variables were examined for significant correlations. These data are presented in Table 8 and Table 9. Given the large number of measured variables, only those that are most relevant to the focus of this study are discussed here. Relevant to this study, it is notable that perceived stereotype threat correlated positively with self-reported change in cognition and dementia worry, as well as with perceived risk of developing dementia both for the self and for other people one’s age. Stereotype threat also correlated negatively with a cognitive composite variable as well as with participants’ ratings of how well they thought they did on the tests. Dementia worry and self-reported cognitive changes also correlated positively with one another. Additionally, dementia worry correlated positively with perceived risk of developing dementia (both for the self and for the average person one’s age), and was associated with a positive family history of dementia as well as with being female. A follow-up t-test demonstrated a significant difference such that women’s mean item response on the DWS (\( M = 1.57, SD = 0.67\)) was significantly higher than men’s (\( M = 1.27, SD = 0.33\)), \( t(89.83) = 3.05, p = .003\), two-tailed (data are adjusted for unequal variance between groups). The difference in mean item response on the DWS between those with a family history of dementia (\( M = 1.61, SD = 0.64\)) and those without (\( M = 1.34, SD = 0.55\)) was also significant, \( t(98.69) = -2.33, p = .02\), two-tailed (data adjusted for unequal variance). In terms of other demographics variables, dementia worry also correlated negatively with age, suggesting participants endorsed less dementia worry the older they were. Correlational data suggested that those who reported more changes in
Table 8

*Correlations Between Cognitive and Demographic Variables*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 CVLT Tot.</td>
<td>—</td>
<td>.76†</td>
<td>.52†</td>
<td>.24*</td>
<td>.29†</td>
<td>.10</td>
<td>.29†</td>
<td>.26†</td>
<td>.28†</td>
<td>.24*</td>
<td>.22*</td>
<td>.24*</td>
<td>.13</td>
<td>.22*</td>
<td>.01</td>
<td>.06</td>
<td>.06</td>
</tr>
<tr>
<td>2 CVLT LDFR</td>
<td>—</td>
<td>.52†</td>
<td>.15</td>
<td>.26†</td>
<td>.06</td>
<td>.20*</td>
<td>.20*</td>
<td>.23*</td>
<td>.26†</td>
<td>.11</td>
<td>.20*</td>
<td>.09</td>
<td>.18</td>
<td>.06</td>
<td>.08</td>
<td>.13</td>
<td></td>
</tr>
<tr>
<td>3 CVLT Rec DI</td>
<td>—</td>
<td>.15</td>
<td>.19</td>
<td>.12</td>
<td>.21*</td>
<td>.08</td>
<td>.06</td>
<td>.22*</td>
<td>.15</td>
<td>.01</td>
<td>.02</td>
<td>.20*</td>
<td>.06</td>
<td>.12</td>
<td>.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 DSF</td>
<td>—</td>
<td>.31†</td>
<td>.23*</td>
<td>.43†</td>
<td>.28†</td>
<td>.39†</td>
<td>.27†</td>
<td>.75†</td>
<td>.14</td>
<td>.01</td>
<td>.18</td>
<td>.12</td>
<td>.02</td>
<td>.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 DSB</td>
<td>—</td>
<td>0.11</td>
<td>.38†</td>
<td>.20*</td>
<td>.31†</td>
<td>.25*</td>
<td>.68†</td>
<td>.12</td>
<td>.14</td>
<td>.05</td>
<td>.12</td>
<td>.03</td>
<td>.20*</td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 TMT Cond. 2&amp;3</td>
<td>—</td>
<td>.55†</td>
<td>.49†</td>
<td>.34*</td>
<td>.48†</td>
<td>.16</td>
<td>.20*</td>
<td>.11</td>
<td>.37†</td>
<td>.12</td>
<td>.09</td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 TMT Cond. 4</td>
<td>—</td>
<td>.44†</td>
<td>.38†</td>
<td>.45†</td>
<td>.45†</td>
<td>.25*</td>
<td>.03</td>
<td>.27†</td>
<td>.01</td>
<td>.02</td>
<td>.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 CWIT Cond. 1</td>
<td>—</td>
<td>.75†</td>
<td>.58†</td>
<td>.14</td>
<td>.01</td>
<td>.06</td>
<td>.29†</td>
<td>.13</td>
<td>.08</td>
<td>.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 CWIT Cond. 2</td>
<td>—</td>
<td>.54†</td>
<td>.26†</td>
<td>.04</td>
<td>.11</td>
<td>.29†</td>
<td>.11</td>
<td>.03</td>
<td>.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 CWIT Cond. 3</td>
<td>—</td>
<td>.19</td>
<td>.10</td>
<td>.16</td>
<td>.24*</td>
<td>.15</td>
<td>.02</td>
<td>.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 RDS</td>
<td>—</td>
<td>0.19</td>
<td>-.01</td>
<td>-.08</td>
<td>.04</td>
<td>.03</td>
<td>.22*</td>
<td>.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 Gender</td>
<td>—</td>
<td>-.13</td>
<td>-.04</td>
<td>.12</td>
<td>-.10</td>
<td>-.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 AD Fam Hx</td>
<td>—</td>
<td>-.13</td>
<td>.15</td>
<td>.12</td>
<td>.41†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 Age</td>
<td>—</td>
<td>.28†</td>
<td>.13</td>
<td>.21*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 Subj. Age</td>
<td>—</td>
<td>-.15</td>
<td>-.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 Old Age Begins</td>
<td>—</td>
<td>-.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 Baseline DW</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* CVLT Tot. = CVLT Trials 1-5 Total Recall; CVLT LDFR = CVLT Long-Delay Free Recall; CVLT Rec DI = CVLT Yes/No Recognition Discriminability (Hits minus False Positives); DSF = Digit Span Forward; DSB = Digit Span Backward; TMT Cond. 2&3 = Mean of TMT Conditions 2 (Number Sequencing) and 3 (Letter Sequencing); TMT Cond. 4 = TMT Condition 4 (Number-Letter Switching); CWIT Cond. 1 = CWIT Condition 1 (Color Naming); CWIT Cond. 2 = CWIT Condition 2 (Word Reading); CWIT Cond. 3 = CWIT Condition 3 (Inhibition); RDS = Reliable Digit Span; AD Fam Hx = Presence/absence of family history of dementia; Subj. Age = subjective age; Old Age Begins = Participant estimate of the age at which “old age” begins. All tests have been coded such that higher scores represent better performance. Gender is coded as 0 = female and 1 = male. AD Fam Hx is coded as 0 = negative family history and 1 = positive family history. The following variables had some missing data: Age (one participant declined to answer); Subj. Age (two participants declined to answer); Old Age Begins (two participants declined to answer).

* p < .05; † p < .01.
Table 9

Correlations Between Subjective Variables, Cognitive Composite Measure, and Demographic Variables

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ECog</td>
<td>—</td>
<td>.37†</td>
<td>.38†</td>
<td>.03</td>
<td>.14</td>
<td>-.30†</td>
<td>.28†</td>
<td>.14</td>
<td>.24*</td>
<td>-.12</td>
<td>.01</td>
<td>.09</td>
<td>-.07</td>
<td>.03</td>
<td>-.14</td>
<td>.17</td>
</tr>
<tr>
<td>2 DWS</td>
<td>—</td>
<td>.23*</td>
<td>.02</td>
<td>.09</td>
<td>-.18</td>
<td>.34†</td>
<td>.24*</td>
<td>.22*</td>
<td>-.06</td>
<td>-.22*</td>
<td>.22*</td>
<td>-.21*</td>
<td>.05</td>
<td>-.06</td>
<td>.57†</td>
<td></td>
</tr>
<tr>
<td>3 ST</td>
<td>—</td>
<td>.08</td>
<td>.17</td>
<td>-.21*</td>
<td>.29†</td>
<td>.28†</td>
<td>.11</td>
<td>-.35†</td>
<td>-.20</td>
<td>.06</td>
<td>.12</td>
<td>.12</td>
<td>.05</td>
<td>.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Effort</td>
<td>—</td>
<td>.46†</td>
<td>-.17</td>
<td>.16</td>
<td>.15</td>
<td>.06</td>
<td>.21*</td>
<td>.06</td>
<td>.18</td>
<td>-.15</td>
<td>.12</td>
<td>-.04</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Difficulty</td>
<td>—</td>
<td>-.45†</td>
<td>.19</td>
<td>.29†</td>
<td>-.03</td>
<td>-.03</td>
<td>.09</td>
<td>.08</td>
<td>.04</td>
<td>.14</td>
<td>-.19</td>
<td>-.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Well</td>
<td>—</td>
<td>-.27†</td>
<td>-.16</td>
<td>-.20*</td>
<td>.18</td>
<td>-.07</td>
<td>-.20*</td>
<td>-.04</td>
<td>-.18</td>
<td>.16</td>
<td>-.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Self AD Risk</td>
<td>—</td>
<td>.68†</td>
<td>.68†</td>
<td>.07</td>
<td>-.22*</td>
<td>.39†</td>
<td>-.16</td>
<td>.16</td>
<td>-.09</td>
<td>.31†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Other AD Risk</td>
<td>—</td>
<td>-.08</td>
<td>.18</td>
<td>-.26†</td>
<td>.20</td>
<td>-.07</td>
<td>.10</td>
<td>-.05</td>
<td>.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 Risk Difference</td>
<td>—</td>
<td>-.08</td>
<td>-.03</td>
<td>.34†</td>
<td>-.14</td>
<td>.12</td>
<td>-.07</td>
<td>.26†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Cog Comp.</td>
<td>—</td>
<td>.07</td>
<td>.13</td>
<td>-.37†</td>
<td>-.08</td>
<td>-.03</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 Gender</td>
<td>—</td>
<td>-.13</td>
<td>.04</td>
<td>.12</td>
<td>-.10</td>
<td>-.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 AD Fam Hx</td>
<td>—</td>
<td>-.13</td>
<td>.15</td>
<td>-.12</td>
<td>.41†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 Age</td>
<td>—</td>
<td>.28†</td>
<td>.13</td>
<td>-.21*</td>
<td>-.15</td>
<td>-.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 Subj. Age</td>
<td>—</td>
<td>.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 Old Age Begins</td>
<td>—</td>
<td>.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 Baseline DW</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. ECog = Everyday Cognition Short Form; DWS = Dementia Worry Scale; ST = perceived stereotype threat; Effort = participants’ self-rated effort; Difficulty = participants’ rating of how difficult tasks were; Well = participants’ rating of how well they believe they did; Self AD Risk = participants’ rating of their perceived risk of developing dementia; Other AD Risk = participants’ rating of perceived risk of developing dementia of average person their age; Risk Difference = Self AD Risk minus Other AD Risk; Cog Comp. = cognitive composite variable; AD Fam Hx = presence or absence of family history of dementia; Subj. Age = subjective age; Old Age Begins = participant estimate of the age at which “old age” begins; Baseline DW = baseline dementia worry, assessed prior to the study. Gender is coded as 0 = female and 1 = male. AD Fam Hx is coded as 0 = negative family history and 1 = positive family history. The following variables had some missing data: Age (one participant declined to answer); Subj. Age (two participants declined to answer); Old Age Begins (two participants declined to answer).

* p < .05, † p < .01
cognitive functioning over the past 10 years rated their performance worse, endorsed higher perceived risk of dementia for themselves, and showed a higher discrepancy between their perceived dementia risk for themselves versus their perceived risk for the average person their age.
CHAPTER IV

DISCUSSION

Effective neuropsychological assessment requires recognizing the potential impact of not only neurological factors, but also psychosocial factors on test performance. One such factor is stereotype threat; particularly for older adults, whom stereotypes often frame as warm but cognitively less competent, concerns related to being evaluated on a stereotyped domain has the potential to negatively impact test performance. Additionally, evaluations of older adults are frequently targeting differential diagnosis between dementia and normal aging. This raises the possibility of a diagnosis threat process through which an older adult who may already worry they are experiencing a dementia process underperforms on testing due to this concern, potentially through a sort of self-defeat manifesting through reduced effort. While the potential effects of age-based stereotype threat and dementia-related diagnosis threat on neuropsychological test performance are concerning of themselves, they may also influence patients’ subjective experiences—including their own evaluation of their cognitive change, their level of anxiety related to developing dementia, and their perceptions of age-related bias against them. The purpose of this study was to examine how age-based stereotype threat and dementia-related diagnosis threat influenced performance on several frequently used neuropsychological tests, as well as how these concepts impacted participants’ self-reported cognitive change, dementia worry, perceived stereotype threat, and effort.
Cognitive Variables

It was hypothesized that stereotype threat and diagnosis threat would comparably impact performance on neuropsychological tests, such that those in these experimental conditions would perform worse on testing than those in a neutral control condition. Evaluated domains included learning and memory, auditory attention and working memory, visual attention and processing speed, and executive functioning. Overall, hypotheses were not supported. Although initial omnibus analysis suggested a small effect of condition, follow-up analyses revealed this effect was driven by performance on a single test of auditory working memory. Additionally, this omnibus effect disappeared when outliers were removed from analyses. Taken together, results suggested no significant effect of age-based stereotype threat or dementia-related diagnosis threat on neuropsychological test performance in this sample. Furthermore, no interaction effects were found for variables proposed to be potential threat moderators, such as education and age. Additionally, controlling for theoretically relevant variables (such as stereotype relevance or task difficulty) did not alter results.

There are several possible reasons no experimental effect was found for cognitive variables. Perhaps of greatest concern is the failure of many participants to accurately recall condition. On a multiple choice question in which participants were asked to indicate the purpose of the study (with the additional instruction to “select the best answer based on the instructions you read”), nearly half of participants selected answers corresponding to a different condition than the one they were in. This improved somewhat in debriefing, when different instructions were described and participants were asked which instruction set sounded like the one they had read. However, accurate
condition endorsement still remained problematic, with roughly 30% of the sample either endorsing the incorrect condition or simply indicating they did not know which instructions they read. It is possible that accuracy was so poor on the multiple choice manipulation check because participants may have either neglected the statement relating study purpose back to the instructions they read, or ignored it altogether and instead chose an option representing what they believed the study purpose to be, regardless of what they read. In either case, it becomes difficult to infer whether or not threat was reliably induced in participants.

Furthermore, threat effects may have even been operational in some control participants who assumed a different purpose for the study than the one provided in their instructions; more generally, participants may have experienced mixed types of threat effects. This possibility is supported by some anecdotal information gleaned from debriefing conversations with participants. In one example, a participant in the control condition mentioned that reading “neuropsychological” brought to mind her mother-in-law with AD. Another control condition participant spontaneously asked during testing whether the study would compare older to younger adults. In another example, a participant in the stereotype threat jokingly remarked, “There’s the Alzheimer’s!” when off by a day in writing the date on the informed consent form (before even being administered the manipulation). Another participant in the diagnosis threat condition described an error as a “senior moment.” In short, clear differentiation between conditions was made difficult not only by questionable recall of which instructions were read, but also by participants’ subjective associations between tasks and both age- and dementia-related topics. Notably, an unpublished dissertation with a similar experimental
manipulation also experienced difficulty with participants “believing” manipulation
instructions, making it difficult to ensure a truly neutral control condition (Page, 2012).

Another possible explanation for a lack of threat effects on cognitive performance
is an ineffective and/or not particularly salient manipulation: perhaps these instructions
were simply not sufficiently threatening for this sample. A lack of potency within the
instructions would also explain why so many participants failed to clearly recall which
set of instructions they read. It is worth noting, however, that instructions were modeled
directly upon those used by Suhr and Gunstad (2002, 2005) in their studies of diagnosis
threat. These instructions were also similar to threat manipulations used in studies of
stereotype threat, such as the work of Abrams and colleagues (2006, 2008; though
notably, these studies also included reference to a planned age comparison) and other
researchers who have induced threat via fabricated newspaper articles that portray
“research findings” on aging and cognition (e.g., Hess et al., 2003; Thomas & Dubois,
2011). That these previous studies have demonstrated threat effects with comparable
manipulations suggests that similar instructions should be sufficient to induce threat
effects, making this perhaps a less likely (or at least, an incomplete) explanation for
findings.

Relatedly, threat manipulations may have failed due to difficulty reliably inducing
stereotype or diagnosis threat effects in a laboratory setting as opposed to a clinical
setting. A research evaluation obviously lacks the high stakes nature of a true clinical
evaluation (in which a diagnosis is anticipated), and may therefore be perceived as a
lower pressure situation by participants. For older adults, this seems somewhat less likely
in the context of stereotype threat given existing literature demonstrating threat effects in
laboratory settings; however, it may explain a lack of hypothesized diagnosis threat results. Worth noting also is that as of yet, diagnosis threat has been considered primarily in the context of mTBI, raising the possibility that such threats do not extend to dementia diagnosis, or may not be induced in the same way. Additionally, the majority of existing age-based stereotype threat research has demonstrated outcomes using research measures rather than the typical tests used by clinical neuropsychologists. Perhaps these clinical measures are less likely to be influenced by stereotype threat effects.

Another possibility is that perhaps there were too many DVs for any manipulation to continue being effective. As noted in Lamont et al. (2015), age-based stereotype threat effects were found to become less pronounced the more distal the measurement from the manipulation. Given the present study required completion of four separate neuropsychological tests (totaling roughly 40 minutes of testing, with multiple subcomponents to each test) in addition to several questionnaires, it is possible that participants were no longer experiencing threat effects after the initial few trials of the first (and longest) neuropsychological test, the CVLT-II. This explanation seems less likely, however, given that no effect of experimental condition was evident on the CVLT-II.

Other possible reasons for a lack of either stereotype threat or diagnosis threat effects on cognitive testing include some specific to the particular sample used. Participants in this study were recruited from the UCCS Research Registry, and many of them were initially part of the UCCS Gerontology Research Database. As such, the majority of participants had participated in previous research with older adults (and indeed, often expected the study to relate to aging). Furthermore, these participants’
familiarity with aging research (and frequent debriefings) may have lent them a different, more positive perspective on aging that was protective from threat effects, in line with the theory of essentialist beliefs about aging (Weiss, 2016). Additionally, research is conducted at the UCCS Gerontology Center, a facility whose waiting area is filled with numerous brochures and other materials related to aging and cognitive health. Creating a truly neutral, non-stereotype-influenced control condition becomes a particular challenge when working with participants who frequently expect to be recruited for age-related research and who are faced with multiple age-related cues upon entering the building, including a younger examiner (particularly considering that presence of a young adult confederate has been used as part of an age-based stereotype threat manipulation in at least one study—Kang & Chasteen, 2009). Furthermore, that many of these participants have completed other studies in the past likely suggests they have become at least somewhat habituated to cognitive tests and the testing environment. (Anecdotally, one participant even remarked on this in debriefing, noting that while she was nervous the first time she participated in a study, she now finds them fun and looks forward to them!) Perhaps more problematic still is the potential for practice effects to influence results, as participants in this study sometimes revealed they had very recently completed another study or other cognitive testing through a nearby clinic that offers annual memory screenings. Although efforts were taken to ensure the precise tests used in this study were not also being used by other researchers or clinics in the area, some tasks (such as trail-making or digit span tasks) are nevertheless quite similar at their core.

Finally, the possibility exists that the present study’s findings represent an authentic lack of effect, and that cognitive effects of age-based stereotype threat and/or
diagnosis threat have been overstated in the literature. With regard to age-based stereotype threat specifically, existing research suggests wide variability in both the magnitude of effects and the variables that contribute to or moderate effects. For example, Lamont et al. (2015)’s meta-analysis observes that effect sizes for dependent variables measured immediately after age-based stereotype threat manipulations range between $d = -4.4$ and $d = 5.5$. Ambiguity of moderator results (for example, competing theories that higher education may increase susceptibility to threat as in Hess, Hinson, & Hodges, 2009, or that lower education increases susceptibility as in Andreoletti & Lachman, 2004) adds further suggestion that much is left to be understood about age-based stereotype threat. With regard to diagnosis threat, although initial studies (Suhr & Gunstad 2002, 2005) found evidence for neuropsychological test performance effects, follow-up studies have had a difficult time replicating these findings. Such studies have instead found some evidence for diagnosis threat effects on subjective variables (e.g., academic self-efficacy, Trontel et al. 2013; memory self-efficacy, Kit et al., 2014; symptom reporting, Ozen & Fernandes, 2011; but see also Blaine et al., 2013 and Carter-Allison et al., 2016 for nonsignificant results on both objective and subjective variables). While the possibility of overstated threat effects is perhaps a troubling reflection of recently-criticized problems in replication in psychological research (for a brief summary, see Pashler & Wagenmakers, 2012), it also offers a glimmer of hope: if threat effects on older adults’ cognitive test performance are not as widespread or severe as some published ABST research would suggest, then there is less cause for concern in terms of clinicians’ inadvertently suppressing patient performance via threat induction.
Subjective Variables

It was hypothesized that those in the stereotype threat condition would have the highest levels of self-reported perceived stereotype threat, that those in the control condition would have the lowest, and that those in the diagnosis threat condition would fall somewhere in between. This hypothesis was supported (although only the difference between the stereotype threat and control conditions was statistically significant). This suggests that despite the previously discussed issues with manipulation checks, reading stereotype-activating instructions nevertheless increased the degree to which participants were likely to endorse perceiving age-related bias against their cognitive functioning. This raises some concern for how neuropsychological testing is presented to older adult patients. Commonly used introductory statements, such as those “reassuring” a patient that their scores will only be compared to people of their own age group (with the attached implication that younger adults’ performances are consistently better), may reinforce a patient’s perception of bias against them. Still, while this effect was significant, mean levels of stereotype threat nevertheless remained low overall in this sample regardless of condition. The highest mean item response by condition was among those in the stereotype threat group in response to the statement, “Based on my age, people often underestimate my cognitive functioning” with a mean item response of 2.58 (on a scale where 1 = strongly disagree and 5 = strongly agree). This suggests that although those in the stereotype threat group reported higher levels of threat statistically, they nevertheless continued to trend toward disagreeing with statements regarding perceptions of age-related cognitive bias (but didn’t disagree as strongly as those in the other conditions).
This raises the question: why weren’t levels of perceived stereotype threat higher? Considering speculation that ageism remains potentially the most prevalent form of prejudice (WHO, 2015) and that the vast majority of Americans over 60 have experienced at least one incident of ageism (Palmore, 2004), it stands to reason that a sample of older adults subjected to an hour’s worth of testing and questions about age-related concerns might feel judged for their age. Anecdotally, participants often expressed agreement (verbally or nonverbally) during debriefing when ageism was described when discussing age-based stereotype threat—making it all the more perplexing that items such as “Some people feel I have less cognitive ability based on my age” were not more frequently endorsed. This could potentially relate to previously mentioned unique qualities of a sample that is used to being recruited for age-related research, and whose perspective on aging has changed as a consequence. It could also reflect our culture’s tendency to overlook ageism as a type of bias—perhaps prejudice is not reported as perceived because it isn’t perceived (e.g., jokes about having “senior moments” may be interpreted not as reflecting prejudice, but rather reality). Another possibility is that to endorse higher levels of perceived stereotype threat would be to acknowledge that one is an “older adult” who can be judged for one’s age. As Barber (2017) notes, those who are perceived as older adults frequently do not see themselves in that way. Results in the present study, where subjective age was almost universally rated as lower than chronological age, and where chronological age was almost universally lower than participant ratings of “At what age does ‘old age’ begin?,” certainly seem to offer at least some support for stereotype distancing as an explanation for globally low perceived stereotype threat ratings.
Hypotheses regarding other subjective variables were not supported. It was predicted that those in the diagnosis threat group would report the most changes in their cognitive functioning as measured by the ECog-12, that those in the control condition would report the fewest, and that those in the stereotype threat condition would fall in between the two. Instead, results suggested that all three groups were comparable in their responses, which on average fell between options characterizing cognition as “better or no change” and “questionable/occasionally worse” over the past 10 years. Thus overall, this sample perceived minimal change in their cognitive functioning, and their perceptions did not appear to be influenced by having their attention drawn to stereotype-based or diagnostic considerations. Globally low ratings of perceived cognitive change may also relate to the possibility of stereotype distancing discussed above—to endorse greater frequency of cognitive slips may have been assumed to be equivalent to identifying oneself as becoming an older adult, and therefore may have been aversive. Alternatively, they may represent accurate assessment of change in a largely high-functioning sample.

Similarly, the hypothesis that those in the diagnosis threat condition would endorse the highest levels of dementia worry, those in the control condition would endorse the lowest, and those in the stereotype threat condition would fall in between was also unsupported. Groups were equivalent in their mean item response on the DWS, and generally endorsed low levels of dementia worry across the board. This suggests that participants’ self-reported anxiety about the possibility of developing dementia was not swayed by being prompted to focus on either age-related stereotypes related to cognition or even dementia diagnosis-related content. Given reports of high levels of fear regarding
development of AD (Harvard School of Public Health & Alzheimer Europe, 2011; MetLife, 2011), low overall endorsement of dementia worry is somewhat surprising. One likely explanation is that while the average individual certainly fears developing AD or another dementia, few older adults experience a fear persistent enough to endorse items included on measures of dementia worry like the DWS (items such as “I find it difficult to control my worries about developing dementia”). While the concept of dementia worry was initially introduced as a broad-spectrum term intended to encompass everything from fleeting concern to genuinely phobic reactions (Kessler et al., 2012), perhaps it would be more useful to differentiate between degree of concern. Those who “cannot stop” worrying about dementia (as in another DWS item) are likely experiencing something fundamentally different from those who simply recognize the horror that a diagnosis of dementia represents. However, existing measures of dementia worry have no published cutoff scores for what represents a normal versus a pathological level of concern.

Finally, the hypothesis that those in the stereotype threat condition would report the highest levels of effort, those in the diagnosis threat condition would report the lowest, and those in the control condition would fall in between was also unsupported. Participants in all three groups reported overall high levels of effort; although mean score on this single item was slightly higher among those in the stereotype threat group, this did not represent a statistically significant difference. While these results represented self-reported effort, effort as assessed via embedded validity measures told a different story. Reliable digit span was significantly lower among those in the diagnosis threat condition as compared to the control condition, and approached a significant difference when compared to the reliable digit span of those in the stereotype threat condition. This offers
some support for literature suggesting that diagnosis threat conditions may lead to reduced effort among participants. However, it is somewhat inconsistent with previous studies’ precise findings using mTBI samples (e.g., Suhr & Gunstad 2002, which found an effect on self-reported effort; Suhr & Gunstad 2005 and Trontel et al. 2013, which found no effect on either self-reported effort or embedded validity measures). A lack of meaningful difference between the age-based stereotype threat and control conditions, while counter to this study’s hypothesis that older adults would be motivated to expend more effort to disconfirm age-related stereotypes, is nevertheless consistent with research suggesting equivalent self-reported effort between threat and non-threat groups (e.g., Steele & Aronson, 1995).

Many of the same issues that plagued assessment of experimental impact on cognitive variables are equally relevant for understanding possible explanations for a lack of experimental effect on the subjective variables of interest. Key among these remains concern regarding participants’ accurate recollection of study instructions. However, here the waters are muddied by the support for the hypothesized relationship between study instructions and stereotype threat. That perceived age-based stereotype threat was significantly higher among those in the stereotype threat condition as compared to those in control offers at least some suggestion that despite questionable levels of accurate recall of instructions, condition nevertheless had some influence—at least for those two conditions. It also calls into question the previously discussed possibility that instructions were not sufficiently threatening to induce effects. Perhaps the manipulation was potent enough to mildly elevate perceived stereotype threat levels in the most vulnerable experimental condition (i.e., the condition whose instructions most aligned with the
construct being measured by the perceived stereotype threat measure), but not to elevate stereotype threat to a level likely to impact other variables of interest (i.e., neuropsychological test performance). This seems supported by the globally low levels of stereotype threat endorsed, even among that experimental group.

A significant experimental effect on perceived stereotype threat effect also raises doubts about distance from manipulation as a possible explanation for a lack of other hypothesized effects. The perceived stereotype threat questionnaire was the penultimate measure (followed only by the demographics questionnaire). If effects were not observed on other measures due to their being too distal from the experimental manipulation (as suggested by Lamont et al., 2015), it seems unlikely that an effect would be observed on a measure administered so late in the study. On the other hand, the stereotype threat measure was administered immediately after the multiple choice manipulation check. This was initially planned to try to minimize participants’ misunderstanding the manipulation check—if they were asked the purpose of the study immediately after answering questions about perceived age-related bias, it was feared they may assume that examining age-related brain changes was the purpose of the study and select this option (regardless of the instructions they read). However, perhaps this placement had the unintended consequence of creating a (mild) threat effect immediately before administering a stereotype threat measure.

Issues related to the testing environment may have also impacted subjective variables. As noted above, recreating the high stakes nature of a clinical evaluation in a laboratory setting is difficult, and could potentially explain why responses were globally low on measures such as the ECog-12. That is, a diagnosis threat effect may be present in
a clinical setting, potentially leading to inflated reports of cognitive change; however, a laboratory setting may not represent enough of a high stakes, threat-inducing environment to elicit such responding (particularly given that participants were explicitly told they would receive no feedback on their cognitive functioning). Additionally, the same concerns regarding overabundance of age- and dementia-related materials available in the research setting hold true for subjective variables as much as cognitive ones.

The same construct-related concerns regarding diagnosis threat considered above in terms of cognitive hypotheses continue to apply when attempting to understand the lack of support for a diagnosis threat effect on subjective variables. Existing diagnosis threat research focuses on participants who have been diagnosed with (or believe they have had) mTBI. Diagnosis threat may not be able to be activated in older adults with no history of dementia diagnosis, and this may explain why instructions aiming to induce diagnosis threat failed to influence subjective variables. Still, research on even mTBI-related diagnosis threat remains sparse and somewhat inconsistent, making it difficult to say whether the present lack of effect on subjective variables represents a true lack of diagnosis threat induction or simply additional evidence of a lack of diagnosis threat effect altogether. For example, while this sample’s lack of increased cognitive symptom reporting is inconsistent with results found by Ozen and Fernandes (2011) using an mTBI sample, other studies (e.g., Blaine et al., 2013; Carter-Allison et al., 2016) have found no support for diagnosis threat effects on symptom reporting.

**Exploratory Results**

Although no evidence of experimental effect was found for subjective age, this study offers additional support for findings that suggest older adults tend to rate their
subjective age lower than their chronological age (e.g., Rubin & Bernsten, 2006). This in itself offers some insight into the generally negative view associated with aging; presumably, participants consider themselves to feel healthier and/or more energetic than the “typical” person of their age group, and therefore assign themselves a lower subjective age. It may also be that participants have not experienced (or are disinclined to report) noticeable changes in their everyday functioning since a given age, and therefore report lower subjective ages. These possibilities are speculative, however, as this study’s data are unable to provide a rationale for why participants responded as they did on a measure of subjective age.

Experimental condition had no effect on either participants’ perceived risk of personally developing dementia or the difference in their risk assessments for themselves versus the average person their age. Viewing the sample as a whole, results suggested that overall participants estimated their own risk as statistically significantly lower than the risk of the average person their age. This is consistent with research on what has been called unrealistic optimism or comparative optimism, which generally finds that people tend to underestimate their own risk of negative events (including health problems) relative to the risk of others (Hevey & French, 2012; Shepperd, Waters, Weinstein, & Klein, 2015). Lower perceived risk of dementia may also relate to results seen in the present study reflecting possible efforts to distance oneself from being viewed as an older adult. It stands to reason that if one does not feel like an older adult, one might also judge one’s risk of age-associated disorders as being lower as well. In any event, a perception of being at lower risk for diseases such as AD and other dementias carries with it the problematic possibility of not seeking testing or treatment until later in the disease course.
That this result was primarily found among participants with no family history of dementia suggests this may be a particular risk for those who (mistakenly) believe a lack of genetic risk factors reflects a lack of dementia risk altogether.

Correlational data provided some interesting insight into relationships between several of the variables of interest within the study. Given the evidence of an experimental effect on perceived stereotype threat, relationships between stereotype threat and other variables are of particular interest. Higher levels of perceived stereotype threat were associated with both higher levels of dementia worry and greater reporting of cognitive changes. This suggests that despite the fact that the manipulation did not work as expected in that most hypotheses were unsupported, some of these variables are related in a way that is consistent with hypotheses. That is to say, clearly there is some relationship between perceptions of age-related bias, harsher judgment of one’s own cognitive changes, and level of anxiety regarding the possible development of dementia. Furthermore, higher reporting of cognitive change and dementia worry also correlated positively. It is somewhat perplexing that stereotype threat would be impacted by the experimental manipulation, yet other variables correlated with threat were unaffected. This may be explained by the fact that among the subjective DV measures, the questions assessing perceived stereotype threat most closely aligned with the content of the experimental instructions. That is, the content of the age-based stereotype threat instructions directly address societal assumptions about aging and cognition in a way that is more personal for an older adult participant, whereas neither the diagnosis threat nor stereotype threat instructions necessarily directly discuss issues related to dementia worry or diagnosis threat. That correlations were significant when manipulation effects were not
may offer evidence that null results are consequences of some of the previously discussed issues with the manipulation (i.e., problem with instruction recall and difficulty with preventing threat effects in control conditions). Alternatively, there may be an additional unmeasured factor (or factors) influencing participants’ perceptions of age-related bias, their thoughts on how their cognitive functioning has changed over 10 years, and their anxiety about developing dementia in the future.

In addition to its positive correlations with stereotype threat and perceived worsening of cognitive functioning, dementia worry also correlated positively with perceived risk of developing dementia. While it stands to reason that those who perceive themselves to be at greater risk of developing dementia also worry more about it as a consequence, it may also be that being inclined to worry more about dementia results in estimating one’s own risk as higher. Results also suggested that within this sample, being younger (relative to the sample), female, and having a positive family history of dementia were each associated with greater levels of dementia worry. Higher levels of worry among those with a family history of dementia is consistent with initial conceptualizations of anticipatory dementia, which considered the concern specifically in the context of adult children of those with AD (Cutler & Hodgson, 1996). These results also provide further support for a growing body of studies that suggest at least among those age 50 and up, greater age is associated with lower levels of dementia worry (Cutler & Brăgaru, 2015; Cutler & Hodgson, 1996; Suhr & Kinkela, 2007). However, the present data are cross-sectional and therefore unable to determine whether this represents a reduction in worry with increasing age, or simply cohort effects. These results also provide some support for a gender difference in levels of dementia worry, a finding that
is consistent with some previous studies in the literature (e.g., Cutler & Hodgson, 1996; Cutler & Hodgson, 2001; Low & Anstey, 2009) but inconsistent with others (e.g., Cutler & Brăgaru, 2015; Kinzer & Suhr, 2015).

Lastly, those who reported more consistent worsening of cognitive functioning over the past 10 years also tended to rate their own performance on cognitive testing more poorly (despite the fact that neither the ECog-12 nor self ratings of performance correlated significantly with a composite measure of cognitive functioning). They also tended to perceive their own risk of dementia as higher, and as more discrepant from their perceptions of risk for the average person their age. Given the additional positive correlation with dementia worry, this suggests some relationship between concerns about dementia and tendency to judge one’s own cognition more harshly, both in an immediate environment (rating performance on just-completed neuropsychological tests) and a longer-term context (rating how cognition has changed over 10 years).

**Study Limitations**

Many study limitations have been considered in the preceding discussion, including problems with accurate instruction recall, difficulty creating “pure” experimental and control conditions, and a sample that was relatively well-practiced in cognitive tests similar to those administered in the present study. In addition to these limitations, the sample consisted exclusively of older adults. Although this was deliberate given this study’s goal of learning more about how age-based stereotype threat and dementia-related diagnosis threat operate in later life, this also necessarily limits generalizability to other samples (e.g., midlife adults, for whom age-related stereotypes are on the horizon but not yet fully relevant). Additionally, the sample was
disproportionately White and female. Thus, it is difficult to know whether the results of this sample would be similar in a more diverse sample (such as one including members of an ethnic minority group, who may have previously experienced other types of stereotype threat effects).

**Directions for Future Research**

Results from this study may have differed had several limitations been anticipated and addressed. The most obvious of these is concern regarding accurate recall of study instructions, and how this may have impacted threat induction. One potential solution to this issue could be taking steps to ensure participants attend more closely to instructions. Rather than simply reading and returning instructions, participants might be asked to write a statement and sign indicating their understanding, or even to copy the instructions completely. Although adding a statement to instructions informing participants that they will be required to remember what they read could confound stereotype threat inductions (some of which simply emphasize the memory-based nature of a task; e.g., Desrichard & Köpetz, 2005 and Rahhal et al., 2001), participants could instead be told they will be asked to select which instructions they read at the end of the test. While the manipulation for the present study was selected to align as closely as possible to Suhr and Gunstad’s (2002) work, modification of instructions might induce a more potent threat effect. For example, the manipulation in this study most closely aligns with what Lamont et al. (2015) described as a “fact-based” manipulation; modification to instead create a more “stereotype-based” manipulation might have achieved different results. Such a manipulation might incorporate reference to planned performance comparisons, such as noting that “Results of older and younger adults will be compared” in the stereotype
threat condition or “Results of those with and without Alzheimer’s disease and other dementias will be compared” in the diagnosis threat condition. An alternative design could include an examiner who is blind to hypotheses but not blind to condition, and could therefore give cues reinforcing experimental condition throughout the study.

In addition to addressing problems with instruction recall, creating a more age-neutral environment for participants could reduce potential confounds in a future similar study. As previously discussed, participants in this study were required to wait in the “UCCS Gerontology Center,” whose waiting area includes numerous age- and dementia-related brochures and signage. (Some participants even brought brochures with them into the otherwise-neutral testing room!) Ideally, a future study such as this one would minimize age-related cues as much as possible. Similarly, participants would ideally not be aware that they are recruited due to their being in a given age group. Furthermore, a more appropriate sample might be one less familiar with the type of research/testing used in the present study.

This raises the question more generally of what represents the ideal sample for a study of this type. Although the present sample was required to be cognitively intact, perhaps this limited the saliency of any diagnosis threat manipulation. While it was presumed for this study that the mere possibility of performing poorly on tests that are often used to diagnose AD and other dementias would be sufficient to induce diagnosis threat, it is possible that this isn’t the case for cognitively intact older adults. This suggests that one avenue for future research would be to use a clinical population. For example, a sample that is self-referred for memory testing may be more susceptible to threat effects due to a pre-existing, as-yet unassuaged concern regarding declines in
cognitive functioning. (Such a population may also be expected to have greater variance on measures such as the ECog-12 and DWS.) Although obviously the precise design of this study would be unethical in a clinical evaluative context, standard parts of a clinical evaluation could be modified slightly to assess their potential threat effects. For example, patients could be randomly assigned to complete a clinical interview (in which a patient is asked to describe in detail the nature of their deficits) either before or after testing as a way of potentially creating a diagnosis threat effect. Or, participants might be randomly assigned to receive information about demographics-corrected scoring before or after testing to assess whether being reminded of assumed age-related changes in cognitive functioning prior to testing (comparable to stereotype threat inductions) impacts functioning. Still another interesting population to explore these research questions with would be those already diagnosed with MCI. This seems the older adult population most comparable to those with mTBI in terms of finding a group for whom a diagnosis is relevant, but not (yet) severe (though assessment of perceived risk of dementia in this sample may be less informative given their objectively higher risk of developing dementia). Repeating a study such as this with a clinical population would add additional relevance for understanding how psychosocial variables such as age-based stereotype threat and dementia-related diagnosis threat might influence the real world work of neuropsychologists.

Despite a lack of support for most hypotheses, correlational evidence for relationships between perceptions of stereotype threat, dementia worry and perceived dementia risk, and self-reports of cognitive change over time suggests that additional research into how these variables are influenced is needed. The present data are unable to
identify causality between these variables; therefore, future research might focus on identifying causal agents in these relationships. For example, it may be that those who perceive greater change in their cognition have higher levels of dementia worry as a consequence of this, or it may instead be that those who worry more about dementia are more hypervigilant to and/or exaggerative of changes in cognition. Longitudinal research measuring each of these variables over time may be able to identify whether change in one variable precedes change in another. Alternatively, experimental manipulations may be devised to alter one variable and measure change in another. For example, participants might be asked to list cognitive changes and then randomly assigned to groups who are told they either reported fewer, the same, or more cognitive changes than the typical person their age, after which dementia worry could be measured. Another potential avenue for future research to understand the correlations found in this study may be to consider other factors that may universally influence such variables as dementia worry, self-reported cognitive change, and perceived stereotype threat. For example, trait-level personality variables such as neuroticism may affect all of these variables and additionally might influence susceptibility to threat-related primes. General health anxiety may also manifest later in life as dementia worry and a proclivity to more harshly judge cognitive functioning and changes in cognition.

Given results suggesting an experimental effect on age-based stereotype threat and evidence of stereotype threat’s correlations with other clinically relevant variables, understanding how common clinical practice (or broader cultural factors) influence older adults’ perceptions of age-related bias and how that bias in turn influences their performance and informs their self-perceptions remains important. Additional studies of
age-based stereotype threat using common neuropsychological measures and more closely approximating the contexts of clinical evaluations are needed to best apply existing stereotype threat theory to a clinical context, where it may have some of the most significant impact on the lives of older adults.

In Conclusion

Understanding the cultural and contextual factors that may influence test performance during a neuropsychological evaluation—and by extension, the results and interpretation of that evaluation—is vital to the ethical practice of neuropsychology. For older adult patients, this requires an understanding of how test performance could be influenced by society’s widespread acceptance of age-related stereotypes, which frequently include assumptions of cognitive decline—raising the risk of age-based stereotype threat and/or dementia-related diagnosis threat effects in cognitive testing. Although the current study did not offer evidence for an experimental effect of the present threat manipulations on neuropsychological test performance, it nevertheless was suggestive of some impact of age-emphasizing instructions on perceived age-based stereotype threat. Furthermore, correlations suggested interrelationships between such factors as perceived stereotype threat, dementia worry, perceived risk of dementia, and self-reports of long-term cognitive change and short-term cognitive performance. Although the possibility remains that age-based stereotype threat and diagnosis threat effects have been overstated in the literature via overemphasis of significant effects and minimization of non-significant or discrepant findings, more research—ideally using multiple alternative methodologies—is needed to ensure these constructs are more thoroughly illuminated, both in general and more specifically in an older adult
population. Additional empirical evidence reflecting a lack of threat effects in older adults may suggest a previously overlooked resilience that warrants further investigation. Given the large proportion of neuropsychological referrals for differential diagnosis of dementia along with the shifting demographics of society toward higher proportions of older adults, psychologists must invest resources in understanding the interaction between culture and cognition.
REFERENCES


Health and Retirement Study (HRS). (2012). Module 6: AD diagnosis and financial planning. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI.


APPENDIX

IRB Approval

University of Colorado
Colorado Springs
Institutional Review Board (IRB) for the Protection of Human Subjects

Date: 7/21/2016

IRB PROTOCOL NO.: 17-094
Protocol Title: A Study of Neuropsychological Tests and Self-Report Questionnaires
Principal Investigator: Joe Madden
Faculty Advisor if Applicable: Molly Marfield
Application: New Application
Type of Review: Expedited 7
Risk Level: No more than Minimal Risk
Renewal Review Level (if changed from original approval) if Applicable: N/A No Change
This Protocol involves a Vulnerable Population: N/A (No Vulnerable Population)
Expires: 20 July 2017

External Funded: ☐ No ☑ Yes
OSP #: Sponsor:

Thank you for submitting your Request for IRB Review. The protocol identified above has been reviewed according to the policies of this institution and the provisions of applicable federal regulations. The review category is noted above, along with the expiration date, if applicable.

Once human participant research has been approved, it is the Principal Investigator's (PI) responsibility to report any changes in research activity related to the project:
- The PI must provide the IRB with all protocol and consent form amendments and revisions.
- The IRB must approve these changes prior to implementation.
- All advertisements recruiting study subjects must also receive prior approval by the IRB.
- The PI must promptly inform the IRB of all unanticipated serious adverse events (within 24 hours). All unanticipated adverse events must be reported to the IRB within 1 week (see 45CFR46.101(c)(2)). Failure to comply with these federally mandated responsibilities may result in suspension or termination of the project.
- Review study with the IRB prior to expiration.
- Notify the IRB when the study is complete.

If you have any questions, please contact Research Compliance Specialist in the Office of Sponsored Programs at 719-255-3903 or irb@uccs.edu

Thank you for your concern about human subject protection issues, and good luck with your research.

Sincerely yours,

Zak Valkyrie

Zak Valkyrie, PhD
IRB Reviewer