Summer 8-15-2015

The Influence of Word Frequency and Aging on Lexical Access

Emily Rebecca Cohen-Shikora
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Psychology

Dissertation Examination Committee:
David A. Balota, Chair
Janet Duchek
Denise Head
Brett Hyde
Rebecca Treiman

The Influence of Word Frequency and Aging on Lexical Access

by
Emily Cohen-Shikora

A dissertation presented to the
Graduate School of Arts & Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2015
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

Endless thanks to my committee members, Dave, Jan, Becky, Denise, and Brett, for embarking on this journey with me. It has been the most empowering and instructive experience of graduate school. A special thanks goes to my core committee, Dave, Jan, and Becky, and to my other academic mentors, Todd, Keith, and Julie. Extended meetings and emails have made your knowledge, expertise, and sheer love for science clear, and for setting those examples for me, I am so grateful. Thank you to the Graduate School and the NIA Training Grant (AG00030) for tuition, stipend, and research support over the years.

Thanks also to staff, faculty, colleagues, and friends here at Wash U. The positive and collegial atmosphere, particularly among my delightful cohort, turned would-be difficulties into common and surmountable academic struggles. Thank you also to the Balota lab, especially Ashley, Mary, Geoff, Jon, and Andy. You were quick with a smile or a gripe session, and generous with your help and support. Thank you St. Louis—fun, baseball, and beer, oh my!—for keeping me happy and, best of all, keeping me.

Thank you to my family, Mom, Dad, Sam, and Evan, whose frank assumptions that I could do all of this was a huge source of motivation. You never believed me when I expressed doubts or insecurities and were incredibly supportive. I only made it this far with your support and love, so thank you! Thanks also to my fur babies, Marley and Archie, who always made me laugh and were marginally less demanding than the dissertation.

Finally, thank you to Brent, the best human being I have ever had the good fortune to know. Your support during the dissertation process is too great to enumerate here but ranged from the large (picking up my slack with dog and house care) to the small (always checking in on me). I can’t wait to be Dr. and Mr. Dulle.

Emily Cohen-Shikora

Washington University in St. Louis

May 2015

ABSTRACT OF THE DISSERTATION

The Influence of Word Frequency and Aging on Lexical Access

by

Emily Cohen-Shikora

Doctor of Philosophy in Psychology

Washington University in St. Louis, 2015

Professor David Balota, Chair

Visual word recognition has been a central area of psychological inquiry over the past century. The current dissertation examines how visual word recognition changes as a function of age by focusing on the influence of word frequency, or how commonly a word is encountered. Word frequency is arguably the strongest predictor of visual word recognition performance across a variety of language tasks, and the most influential factor in models of language processing. All models of visual word recognition include a strong role for word frequency but often assume different underlying mechanisms, which produce differing predictions for age changes. Although there is already a literature examining word frequency effects in younger and older adults, these studies have produced inconsistent results, possibly due to procedural limitations and task-specific processes. This dissertation explores the influence of task and age on the word frequency effect, while directly examining individual differences (e.g., changes in vocabulary, vision, education) in order to better understand the mechanisms underlying word frequency effects. In contrast to the dichotomous approach of examining extreme groups of young and older adults, or extreme bands of word frequency, the present study examined both variables in a continuous manner. The primary finding is that the word frequency effect does not appear to change as a

function of age across all three tasks considered. This finding is discussed in reference to previous inconsistent findings in the literature and important theoretical implications.

Chapter 1: Introduction

Language is perhaps the most critical of higher-level cognitive skills. It is a distinctly human ability which every modern and ancient community has expertly used for communication. Not only is language an important skill that has been studied in isolation, it is also a complex building block of other cognitive skills, and is used to assess other higher-order cognitive functions. Although older adults often display deficits in these other higher-order abilities such as memory (e.g., Craik, 1994), attention (e.g., Craik & Byrd, 1982; Hasher & Zacks, 1988; Madden, 2007), and decision-making (e.g., Boyle et al., 2012), there is less prominent age-related decline in language performance, with older adults even outperforming young adults in some cases (e.g., Verhaeghen & Salthouse, 1997).

The current dissertation will address age-related changes in the visual word recognition process. Some aspects of visual word recognition tasks, and language more broadly, show age-related deficits. For example, encoding input, producing output, and executing motor responses in a task are all particularly susceptible to age-related decline (Balota & Duchek, 1988; Ratcliff, Thapar, Gomez, & McKoon, 2004), and word-finding difficulties increase in older age (Abrams & Farrell, 2011; Lovelace & Twohig, 1990; Ossher, Flegal, & Lustig, 2012). However, there is also preservation in the representations themselves, as reflected in semantic memory tasks (Balota & Duchek, 1988, 1991; Burke & Shafto, 2008). Additionally, older adults often show substantially higher vocabulary knowledge than young adults (Verhaeghen, 2003), and their performance in tasks assessing processing speed and working memory may show less decline when stimulus materials are verbal rather than spatial (Jenkins, Myerson, Joerding, & Hale, 2000).

1.1 Age Differences in Language Processing

Older adults' language processing skills are more developed and, at the same time, more subject to age-related insult than younger adults'. Older adults have used language for several decades longer than the typical twenty-year old and are often more highly educated (Verhaeghen, 2003). Because of this, it is tempting to view older adults' preserved language performance as a consequence of increased practice or repetition across the lifespan (e.g., Gollan, Montoya, Cera, & Sandoval, 2008). There is certainly some contribution of increased practice and repetition to age effects; however, this is not all that aging involves. There is evidence that aging is associated with both widespread decline and specific breakdowns that may influence lexical processing, most notably within speed of processing, sensory systems, and attentional control. The goal of the current project is to examine visual word recognition in adults across the age spectrum, considering the factors that typically change with respect to age: experience-related benefits and age-related declines in component systems.

Older adults' greater experience with language involves more exposures to each word, richer semantic network connections, and more encounters with each word in different contexts. Younger adults with greater language skill and experience show evidence of more accurate, efficient, and even automatic word processing, as indexed by a smaller influence of standard predictor variables on performance (e.g., neighborhood structure, word frequency, semantics, etc.; Yap, Balota, Sibley, & Ratcliff, 2012), relative to less skilled and experienced younger adults. The natural extension of this work is the examination of older adults, who have more reading experience, and often more skill than young adults (e.g., higher vocabulary knowledge, Verhaeghen, 2003).

There are a number of domains which show age-related breakdowns, including processing speed (Salthouse, 1996), sensation and perception (Fozard & Gordon-Salant, 2001), and attentional control (Kramer & Kray, 2006). Processing speed differences are a notorious confound in aging research, particularly when response times (RTs) are the dependent variable of interest, and processing speed often accounts for much, if not all, age-related variance in a cognitive task (e.g., Salthouse, 1994, 1996). Similarly, vision is implicated in visual word recognition tasks but also declines with increasing age (Schieber, 2006). A potential consequence of such sensory decline is that visual information coming into the system is partially degraded, and older adults must engage cognitive effort to decode the stimulus. This expenditure of cognitive effort because of sensory decline has been shown to impede additional higher-level processing such as comprehension or memory (Cronin-Golomb, Gilmore, Neargarder, Morrison, & Laudate, 2007; McCoy, Tun, Cox, Colangelo, Stewart, & Wingfield, 2005; Tun, McCoy, & Wingfield, 2009; Wingfield, Tun, & McCoy, 2005). Thus, sensory or perceptual ability can also act as a major mediator of age-related change. Aging is also associated with declines in attentional control and working memory, abilities thought to be important for even the lowest-level language processing task (Hasher & Zacks, 1988; Salthouse & Babcock, 1991). There is also evidence that this aspect of cognition accounts for age-related differences in language processing; some studies show a direct relation between attentional measures and language performance (Carpenter, Miyake, & Just, 1994; Norman, Kemper, & Kynette, 1992; Sommers & Danielson, 1999; Van der Linden et al., 1999; although see Waters & Caplan, 1996; Wingfield, Waters, & Tun, 1998). The current dissertation assesses the relative influences of processing speed, vision, and working memory in visual word recognition.

Of course, one might argue that automatic processes involved in processing individual words may be impervious to age-related change in processing speed, sensation and perception, and attention. However, it is also possible that visual word recognition will be sensitive to changes in these processes. Because of this, one must consider the separate, and potentially interacting, influences of these factors on older adults' performance. The current dissertation considers these factors; studying multiple underlying components also allows one to better localize any observed age-related performance changes.

1.2 Methods of Studying Visual Word Recognition

Although there is a considerable literature on connected discourse processing during reading (see for example Rayner, Pollatsek, Ashby, and Clifton, 2011), the vast majority of studies of visual word recognition have involved relatively simple tasks such as naming, lexical decision, and semantic categorization. For example, in the naming task, participants are presented with a word (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995; Yap & Balota, 2009) which they must name aloud. Another common task is lexical decision (Balota & Chumbley, 1984; Ratcliff et al., 2004), in which participants are presented with a letter string and must indicate whether the stimulus is a word or a nonword (e.g., "flirp"). In semantic categorization, participants must perform a decision involving the meaning of the word; for example, whether it is living or not (animacy judgment, as in Hino, Lupker, & Pexman, 2002) , or whether it belongs to a given category (category verification, as in Balota & Chumbley, 1984). These tasks have yielded an enormous amount of information on the process of recognizing words (see Yap & Balota, 2015, for a recent review).

In addition to providing information about visual word recognition, tasks also encourage some task-specific processing (Balota & Chumbley, 1985). Word naming, for example, emphasizes grapheme-to-phoneme conversion, and it typically shows strong influences of phonological word onsets, phonological neighborhood (number of similar-sounding words, or average distance to phonological neighbors), and regularity of the grapheme-to-phoneme correspondences (Balota & Yap, 2006), whereas lexical decision shows stronger influences of measures tapping semantic variables (e.g., meaningfulness, imageability, and familiarity, Colombo, Pasini, & Balota, 2006; Hargreaves & Pexman, 2012; Yap, Pexman, Wellsby, Hargreaves, & Huff, 2012), because these dimensions are diagnostic of words but not nonwords. Animacy judgment also involves semantics, as well as lexical variables such as word frequency (e.g., Andrews & Heathcote, 2001).

The current dissertation affords an opportunity to explore variables that are common across tasks and variables that may have task-specific influences. The influence of age will be particularly interesting because there is some suggestion (in Balota et al., 2004) that older adults engage in less task-specific processing compared to younger adults. This is based on the observation that the correlation across lexical decision and naming is higher for older adults than for younger adults.

1.3 Theories of Visual Word Recognition

In addition to considering visual word recognition across the adult age spectrum in relation to lower-level cognitive components of aging such as general slowing and sensory and attentional decline, it is important to consider what studying an aging population may contribute to current models of word recognition. Interestingly, the current models predict differing effects of both experience and age, so the study of visual word recognition across the age spectrum affords a unique opportunity for adjudicating among the relevant models. Although many interesting issues arise with respect to the aging visual word recognition system, in order to keep the dissertation project to manageable scope, the question of interest focused on word frequency. Word frequency is a metric of how often a word is encountered in the language, and is arguably the most robust predictor of RT across a range of experimental tasks. Because of this, all models and theories of visual word recognition account for the word frequency *effect*, i.e., faster RTs and higher accuracy for higher-frequency words relative to lower-frequency words. Models of visual word recognition have different mechanisms that account for the word frequency effect and hence make different predictions about how word frequency effects may change with age. Furthermore, some theories consider age specifically (e.g., Spieler & Balota, 2000), whereas other theories implicitly equate age with simple increased experience. The current dissertation assesses the extent to which age-related experience is implicated in visual word recognition, and examines additional age-related and age-independent factors of interest.

The logogen model (Morton, 1969), posits that word frequency has its effect in reducing the threshold for activation of word units. The benefit of increased exposure approaches an asymptote, at which point further frequency of use does not produce an added benefit for lexical access. This puts a ceiling on the benefit afforded to high-frequency words and, with increased experience, allows low-frequency words to "catch up" to the performance levels of higher-frequency words. Thus with the experience garnered with age, one might predict that lexical access of words across the entire frequency range would approach asymptotic RT performance (see Murray & Forster, 2004, for some discussion of these predictions). Hence, without any embellishment or consideration of additional age-related factors, this model predicts a reduction in the word frequency effect as a function of age.

The transmission deficit hypothesis (TDH; Burke & MacKay, 1997), on the other hand, suggests that aging involves weakened transmission of activation across the entire lexical system, with disproportionate detriments to performance when the target connections are infrequently accessed. As Burke and Shafto (2008) point out, "Using the transmission deficit model, James and MacKay [2001] argue that frequent and recent use of high frequency words maintains the strong connections in their representations, aiding their retrieval. Connections for low frequency words, however, weaken from disuse and from aging which both cause transmission deficits that impair retrieval." (p.407). The TDH therefore predicts larger word frequency effects for older adults than younger adults.

Another prediction regarding age and word frequency comes from the rank frequency account (as in Murray & Forster, 2004). This framework organizes the lexicon into frequency-ordered bins, and considers *relative* word frequency (a word's position within the spectrum of frequency of use) rather than *absolute* word frequency (exact number of encounters with a word). Rank frequency theory predicts equal word frequency effects across age groups, since the relative rank frequencies of words should not change with the increased exposure to words that older adults have. In fact, Murray and Forster cite findings of age-constancy in word frequency effects as evidence for their rank hypothesis and against a more absolute frequency idea such as Morton's (1969) logogen model, in which each absolute encounter with a word changes threshold (Murray & Forster, 2004, see page 724).

Spieler and Balota (2000) put forth another theoretical account, specific to the process of word naming, concerning the word frequency effect and aging. Their account was based on research suggesting that children's lexical access and representations become more unitized and less piecemeal over time; that is, less likely to be processed by being broken down by sub-components like letters and letter pairs and more likely to be processed as an whole-word entire unit (Samuels, LaBerge, & Bremer, 1978). Spieler and Balota (2000) proposed that older adults may show even more unitization and holistic processing, as a continuation of the process of becoming a skilled reader. They measured unitization/holistic processing in terms of word frequency—a larger word frequency effect in an older adult group relative to college-aged young adults was taken as evidence for greater use of direct, whole-word holistic lexical access. Allen and colleagues (Allen, Bucur, Grabbe, Work, & Madden, 2011; Allen, Madden, & Crozier, 1991; Allen, Wallace, & Weber, 1995) posit a similar interpretation of the word frequency effect as suggestive of increased holistic processing, but find equal word frequency effects across age across several visual word recognition tasks. They therefore conclude that holistic processing is similar across age, and instead point to disruptions in part or sub-component processing. Therefore both of these theories might similarly accommodate increased word frequency effects with age by appealing to qualitatively different processing strategies engaged in by older adults.

Another set of theoretical frameworks which would predict increasing word frequency effects with age as a result of a qualitatively different lexical processing strategy come from the eye tracking literature. On the basis of some empirical work (Kliegl, Grabner, Rolfs, and Engbert, 2004; Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006), both the SWIFT model (Laubrock, Kliegl, & Engbert, 2006) and the E-Z Reader model (Rayner et al., 2006) capture aging effects by appealing to a "partial reading" strategy, in which older adults are more likely to guess what a word is on the basis of partial information. This is presumably due to older adults' declining sensory processing and increasing language experience relative to young adults, and the fact that the task involves word recognition in a sentence, rather than isolated word, context. Hence the SWIFT and E-Z Reader models predict larger word frequency effects in older adults compared to younger adults.

In summary, the logogen model predicts a smaller word frequency effect for older adults; the TDH, unitization/holistic processing accounts, and SWIFT and E-Z Reader models predict a larger word frequency effect for older adults, and the rank frequency account predicts no difference in the word frequency effect across age groups. These models and theories vary in their consideration of age and experience, and the specific tasks for which they have explicit predictions. Therefore, studying the word frequency effect across the adult age spectrum affords considerable potential for understanding how word frequency modulates visual word processing.

1.4 Prior Work on the Age by Word Frequency Interaction

Because there are important reasons to model and explore the aging visual word recognition system, it is important to review the somewhat inconsistent literature. As Balota, Yap, and Cortese (2006) note, "understanding the operations in the tasks used to build models of word recognition is a paramount first step in building adequate models" (p.315). Unfortunately, there is little consensus on the ways in which visual word recognition does or does not change with increasing age. As discussed below, there is evidence for larger, equivalent, and smaller word frequency effects in young versus older adults.

One of first few studies to examine visual word recognition in an aging population found equivalent word frequency effects in young and older adults. Specifically, Bowles and Poon (1981) examined the age by word frequency interaction in a double lexical decision task, in which participants were instructed to make one response when both of the presented letter strings were words, and a different response otherwise. The Age by Word Frequency interaction was not significant, although older adults showed a numerically larger word frequency effect (422 ms, as compared to 314 ms in young adults). This lack of interaction persisted in an analysis attempting to control overall processing speed differences between younger and older adults which used response time on a two-choice task as a covariate. Another early study from Tainturier, Tremblay, and Lecours (1989) used a more standard lexical decision task in which only one word (in French) was presented for a word/nonword decision. They, too, did not find an age difference in the word frequency effect, as measured by comparing regression coefficients for word frequency (considered continuously) on raw response times for the younger and older group separately, and by subjecting the data to an analysis of variance (ANOVA) comparing the difference between lowest and highest frequency items for younger and older adults. Their study was also innovative in that they equated education level across age groups. This is important in light of their later study (Tainturier et al., 1989), in which they found a negative association between word frequency effects and educational level. In fact, controlling for or measuring demographic variables which have potential influences on the word frequency effect (e.g., years of education, vocabulary knowledge) marks the introduction of a critical consideration in aging work; younger adults typically have fewer years of education and lower performance on vocabulary measures than older adults (Verhaeghen, 2003). However, not all studies following Tainturier et al. (1989) have similarly controlled for, or investigated, the influence of demographic variables.

Following Tainturier et al. (1989), Allen and colleagues (Allen et al., 1991; Allen, Madden, Weber, & Groth, 1993) examined the word frequency effect as an indicator of holistic processing since a whole-word lookup would be greatly affected by prior frequency of use. They also examined surface manipulations of the stimuli (e.g., letter case mixing, word spacing) to examine detailed, part-word visual word recognition. The results showed a greater impact of surface manipulations on the older adults than on the young adults, but not the hypothesized increasing word frequency effect with age; instead, word frequency effects were constant across age group, even when raw response times were transformed to account for processing differences between younger and older adults (in Allen et al., 1993, only). They therefore concluded that lexical processing did not increase with age, but detailed processing decreased. However, some caveats to this set of studies potentially detract from their conclusions. For example, in Allen et al. (1991), older adults showed a numerically larger word frequency effect (53-ms larger than young adults), and in Allen et al. (1993), word frequency effects were, in general, smaller than is typically found (no overall main effect of word frequency). So it seems that these studies are limited in their interpretability. Importantly, in these studies, older adults showed higher vocabulary than younger adults.

Later studies by Balota and Ferraro used a word naming task (Balota & Ferraro, 1993) and a lexical decision task (Balota & Ferraro, 1996) to examine the word frequency effect in younger adults, healthy older adults, and older adults with varying levels of Alzheimer's disease. Frequency effects in these studies increased as a function of age and dementia status. Importantly, in the Balota and Ferraro (1996) paper, vocabulary was equated across the younger and older adults, and word frequency effects were specific not associated with overall response latencies or accompanied by increases in the effects of all variables from the younger to the older age group, meaning that general slowing across the lifespan cannot account for the patterns of results. Similarly, Spieler and Balota (2000) conducted a large-scale word naming task and found an age-related increase in the word frequency effect, as measured by word frequency regression coefficients for the younger versus older adults.

Another recent set of studies employed eye tracking, in which participants engaged in a text reading task while their eye movements and fixations were measured. Kliegl and colleagues (2004) and Rayner and colleagues (2006) examined target words varying in word frequency values and language (two levels of frequency in English in Rayner et al., 2006; five levels of frequency in German in Kliegl et al., 2004), embedded within neutral sentences. In both studies, older adults showed a larger influence of frequency across several measures of eye movement (but similar effects of other variables, suggesting they are not due to overall slowing across the lifespan). Older adults also showed more word skipping and regressions back to a word when reading a text (Kliegl et al., 2004; Rayner et al., 2006), which is supportive of the proposition that they were relying more on a guess-based partial reading strategy, in which they decode only part of the word (Laubrock et al., 2006; Rayner et al., 2006). This is presumably due to older adults declining sensory processing and more language experience, relative to young adults. However, it should be noted that another eye tracking study found similar word frequency effects between the age groups when stimuli were matched on length and parts of sentences were masked during reading (Rayner, Yang, Castelhano, & Liversedge, 2011). Younger and older adults were equated on vocabulary performance in Laubrock et al. (2006) and Kliegl et al. (2004), but no other studies measured vocabulary.

There have been a few studies of aging and word frequency that used a different task: picture naming. Thomas, Fozard, and Waugh (1977) found no change in the word frequency effect in RTs as a function of age. However, their stimuli were repeated in eight successive blocks of trials and analyses were only conducted on all blocks. In light of the robust finding that frequency effects decline, or are eliminated, with repetition (Balota & Ferraro, 1996), interpretation of Thomas and colleagues' results in the context of the current goals is limited. Newman and German (2005) also used a picture naming task, but their dependent variable was accuracy. They found no significant difference in the frequency effect with increasing age. An unpublished study by Chae, Burke, and Ketron, (2002), mentioned in Gollan et al. (2008), found an increasing word frequency effect with age, but in only raw RT (which doesn't control for age-related general slowing). Finally, Gollan and colleagues examined picture naming in mono- and bilingual young and older adults, and found only non-significant trends towards larger word frequency effects for older monolinguals and bilinguals (relative to younger monolinguals and bilinguals) in English and smaller word frequency effects for older bilinguals (relative to younger bilinguals) in Spanish, even after accounting for age-related slowing. This is particularly interesting because Gollan and colleagues' (Gollan et al., 2011, 2008) theory of the bilingual disadvantage in visual word recognition hinges on less cumulative use for each word in the lexicon. That is, they speculate that the reason one typically sees slower and less accurate performance on visual word recognition tasks for bilinguals than for monolinguals is because their frequency of use for each word is split across two lexicons, one for each language they speak. Functionally, the result is that each word has been experienced fewer times by a bilingual than a monolingual. Gollan and colleagues draw a direct parallel between cumulative experience differences between mono- and bilingual adults, and young and older adults. They therefore interpret larger frequency effects for bilinguals (as found in their first experiment) and for young adults (found in their second experiment as a trend in Spanish only) as evidence for their theory because they are both groups with less cumulative experience. Thus examining the interaction between age and word frequency is critical to this account as well.

Only one study of aging and word frequency included more than one task: Balota et al. (2004). In this large-scale study, younger and older adults did naming and lexical decision tasks with over 2,800 words which varied continuously along the word frequency spectrum. This study examined word frequency regression coefficients as a function of age using several different frequency measures, but the general pattern was that older adults showed a slightly larger word frequency effect for the naming task (measured as change in R^2 value when word frequency was added to the regression model), and the young adults showed a larger word frequency effect for the lexical decision task. This highlights the need to consider task-specific processing, which is a major aim of the current dissertation.

1.5 Inconsistencies and Differences among Studies: Motivation for the Current Dissertation

Because there has been such variation in design, methodology, participants, and stimuli in previous studies, it is not altogether surprising that there is little consensus on whether (and if so, how) the word frequency effect changes with age. It was therefore the goal of the current study to address these issues by conducting a large-scale, comprehensive examination of the age by word frequency interaction. Considered next are the factors which differ among the previous studies, which are likely to contribute to the disparate results. This section will be followed by a section detailing how the dissertation addresses each of these factors.

As described above, prior literature on aging and the word frequency effect has used a diversity of tasks. Most theoretical accounts lack consideration of task-specific influences and few studies employ more than one task. The current study employs three standard visual word recognition tasks, included word naming, lexical decision, and animacy judgment (semantic categorization). The use of multiple tasks allows for triangulation of both task-general (such as general lexical processing or semantic memory) and task-specific (such as phonological or semantic processing) influences. This procedure affords comparison to prior work using the selected tasks.

A second set of issues in prior work concerns participant characteristics, which vary greatly across the studies. Studies often use an extreme-groups design in which college students represent the young population and community-dwelling older adults from a volunteer pool represent the older population. There are several potential problems with this (described in Salthouse, 2000). Briefly, an extreme-groups design may decrease power (Cohen, 1983), and assumes an intermediate middle-aged group (which may not be the case, since the multiple influences on processing such as deficits associated with age and greater experience associated with age may unfold on different time scales). Further, young and older participants typically differ on relevant baseline measures such as education, vocabulary, and attention (Craik & Byrd, 1982; Verhaeghen, 2003). Some studies control for these differences (e.g., Balota & Ferraro, 1996; Gollan et al., 2008; Tainturier et al., 1989), others measure and make note of them (e.g., Allen et al., 1991, 1993), and still others account for variables other than vocabulary (e.g., Balota & Ferraro, 1993), potentially jeopardizing the comparability of results across studies. Finally, a particularly important difference between young and older adults groups are differences in overall response latencies, which will naturally lead to older adults producing larger absolute frequency effects when measured in raw RTs. Thus one must examine word frequency effects above and beyond any general slowing, and indeed there are a variety of ways to do this (e.g., using *z*-scores of RT instead of raw RT, examining word frequency effects as proportions of overall RT, using overall RT as a covariate). The current study recruited a diverse group of participants across the adult age spectrum from a community volunteer sample. Relevant sensory, cognitive, experience, and ability changes that are often correlated with age or the word frequency effect were also measured. Furthermore, age-related differences in processing speed were controlled for by using *z*-scores in most analyses.

A third set of problems for studies of visual word recognition is with stimulus selection and use of appropriate norms. When using words as the target unit, there are many highly correlated variables that must be accounted for. For example, low- and high-frequency words differ on other dimensions than just frequency (e.g., length, consistency, etc.). Accounting for these correlated variables requires extensive matching in a factorial design or many observations to have the power required to include all of the potential third variables in a regression model (Balota, Yap, Hutchison, & Cortese, 2012). Further, there are considerable problems with stimulus selection when attempting to match variables in a factorial design. Forster (2000) noted that experimenters are adept at identifying which of two words might have a faster RT, even when they are matched on certain important characteristics. This may, implicitly or otherwise, influence stimulus choices in an experiment. Previous studies differ as to the extent to which they control for extraneous variables. For example, Balota et al. (2004) considered frequency continuously and used a regression analysis with many other variables in the model to control for their effects; Kliegl et al. (2004) and Rayner et al. (2011) matched the word frequency groups on length, and Morrison, Hirsch, Chappell, and Ellis (2002) matched the word frequency groups on age of acquisition and length. These differences may jeopardize comparability across these studies. Furthermore, a thorough analysis requires stimuli that range along the word frequency spectrum rather than only those at the extreme ends of the continuum. This is particularly important because there is no consistently used boundary for “low” and “high” frequency words, so selection of words for frequency groups varies greatly among studies. Small, artificially dichotomized stimuli sets may yield spurious interactions due to inflated type I error rate and overestimation of effect sizes (Cohen, 1983; Unsworth, Redick, McMillan, Hambrick, Kane, & Engle, 2015), preclude examination of more than one related variable at a time, potentially confound correlated variables, and could be subject to list context effects (described in Balota et al., 2004). The current dissertation avoided these concerns by treating word frequency, like age, as continuous and collecting enough observations for each participant and each item to have sufficient power for a regression analysis with many predictor variables (although collinearity can still be an issue, see for example footnote on p. 18). Finally, another issue related to stimuli is that prior studies have used different frequency metrics. Most of the earlier studies have used Kučera and Francis's (1967) word frequency norms, which are based on a small, outdated corpora and have demonstrably lower utility (see Balota et al., 2004; Brysbaert & New, 2009) than more modern word frequency norms based on larger corpora, such as Hyperspace Analogue to Language (HAL; Lund & Burgess, 1996), the Educator's Word Frequency Guide (Zeno; Zeno, Ivens, Millard, & Duvvuri, 1995), and Subtitle Frequency (SUBTLEX; Brysbaert & New, 2009). It is also possible that there are cohort effects because older word frequency norms may be more consistent with older adults' frequency of exposure than younger adults' frequency of exposure (see Balota et al., 2004). The current study directly examines the influence of frequency norm used on younger and older adults' performance, and uses the most robust metric, SUBTLEX, across analyses.

1.6 Item Predictor Variables in the Current Study

The primary interest in the current study is whether and, if so, how the word frequency effect changes with age. Because of the multicolinearity of word variables, many of which are correlated with word frequency, addressing this question requires consideration of a full set of predictor variables. The following section delineates the variables considered, some typical observations, and their theoretical importance.

Step 1

Phonological Onsets. Phonological onsets are a critical variable to consider in tasks using a voicekey and microphone, as in the current naming task. Different word onsets may trigger the microphone at differential sensitivity (e.g., a hard "k" sound may trigger it more easily or rapidly than a soft "h" sound). In order to address this, a set of 13 dummy-coded variables were included to represent the features of initial phonological onsets (absence or presence of: voicing, bilabial, labiodental, dental, alveolar, palatal, velar, glottal, stop, fricative, affricate, nasal, and liquid; see Kessler, Treiman, & Mullennix, 2002; Treiman et al., 1995). Phonological onsets have been shown to predict as much as 35% of variance in word naming (Balota, Yap, Cortese, Hutchison, Kessler, Loftis, et al., 2007), and should explain relatively little variance in lexical decision or animacy judgments since these tasks do not demand a voicekey.

Step 2

Orthographic Length[1]. Length was measured in the current study as number of letters (versus, for example, number of syllables, bigrams, or phonemes). Orthographic length has been contentious in the literature, in part because effects of length on RT have been traditionally thought to reflect serial, left-to-right processing (e.g., Weekes, 1997) which is a controversial aspect of visual word recognition models (but see Plaut, McClelland, Seidenberg, & Patterson, 1996, for how length effects may be implemented in a model which does not assume serial processing, or see Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001, and Whitney & Cornelissen, 2008, for some discussion of implemented serial processing showing null or reversed length effects). Length effects are also inconsistent across prior studies, but in the most comprehensive lexical decision study in the English language, New, Ferrand, Pallier, and Brysbaert (2006) found a linearly increasing function relating raw RT to length, but a U-shaped function relating length to RTs with the effect of other variables partialled out (e.g., word frequency, orthographic N). The U-shaped function reflected facilitatory effects of length for

[1] In a set of preliminary analyses, orthographic N and phonological N (Coltheart, Davelaar, Jonasson, & Besner, 1977), and orthographic Levenshtein distance and phonological Levenshtein distance (Yarkoni, Balota, & Yap, 2008) were considered as well. However, because these variables are highly correlated with length ($r = .63$ to $r = .86$), the addition of these variables did not seem to add any unique information. This makes it is likely that part of the length effect includes a neighborhood effect in the present study. Importantly, the inclusion of neighborhood effects did not change the influence of age and word frequency in any of the tasks.

very short words, inhibitory effects for very long words, with a null effect of length in the middle 5-8 letter range (replicated in lexical decision and naming by Yap & Balota, 2009). These findings highlight the strength of using a wide range of stimuli, since selecting stimuli with restricted length and partialling out other variables may lead to disparate findings. Length effects are also typically larger in naming than in lexical decision, at least for response latencies (Balota et al., 2004; Yap & Balota, 2009).

Word Frequency. As described in some detail in the introduction, word frequency is the primary variable of interest. In the current study, SUBTLEX word frequency was used in the primary analyses because previous evidence indicates that this is the measure that is most predictive of lexical decision and pronunciation performance (see Brysbaert & New, 2009). This metric was calculated from movie and television subtitles from the years 1900-2007 and included 51 million words total.

Concordance. Concordance is based on a norming study described in more detail below. It was computed on the basis of animacy ratings (e.g., participant ratings of “definitely non-living”, “mostly non-living”, “ambiguous”, “mostly living”, or “definitely living”). Concordance values were defined as the number of people who rated the word as “definitely” or “mostly” living or non-living, divided by the total number of ratings for that word. For example, the word “avocado” received 11 ratings of “definitely living”, 6 ratings of “mostly living”, 1 “mostly non-living”, and 4 “definitely non-living” out of ratings by 22 different participants, and hence it was categorized as living with a concordance score of (11 + 6) / 22 = .77. Concordance was expected to influence animacy judgments, but was not predicted to influence the other tasks except to the extent that it is correlated with other predictor variables.

Feedforward Rime Consistency. Feedforward consistency refers to the extent to which a word's pronunciation is like similarly-spelled words. For example, a word like "spook" is low in feedforward rime consistency since it is not pronounced like similarly spelled words, e.g., "book", "nook", "took", "look", "rook", etc. This variable was computed (based on Yap & Balota, 2009) by calculating how often the word was pronounced like similarly spelled words separately for the rime of each syllable and averaging these values for each syllable (up to three syllables). This measure of consistency did not take into account the frequency of the similarly-spelled words (i.e., was a type measure, not a token measure). Consistency effects are typically facilitatory in naming (Jared, McRae, & Seidenberg, 1990) and have even been found in lexical decision in some studies (Andrews, 1982; Balota et al., 2004), suggesting at least some activation of phonology in visual word recognition. Feed*back* consistency and onset consistency have also been explored, but in the current study only feedforward rime consistency was considered because it is the most reliable in prior literature (Balota et al., 2004; Kessler, Treiman, & Mullennix, 2008; Ziegler, Petrova, & Ferrand, 2008).

Valence. Valence is an affective variable reflecting pleasantness as rated by participants from 1 (unhappy) to 9 (happy) (Warriner, Kuperman, & Brysbaert, 2013). Valence, and affect more generally, have been of recent theoretical interest. Prior research has shown that valence affects naming and lexical decision such that negative words produce slower response latencies (Algom, Chajut, & Lev, 2004; Estes & Adelman, 2008a, 2008b, but see Larsen, Mercer, Balota, & Strube, 2008). This finding is taken as evidence for automatic vigilance, and subsequent slowed disengagement, for negative stimuli (but not positive stimuli, see Estes & Adelman, 2008b). The current study examined the effect of valence with the prediction that negative words should be named more slowly than positive words. Indeed, based on socio-emotional theory by

Carstensen and colleagues (e.g., Mather & Carstensen, 2005), one might even expect older adults to show this effect more strongly, as they typically show a positivity bias relative to younger adults across several cognitive domains.

Concreteness. Concreteness (Brysbaert, Warriner, & Kuperman, 2014; James, 1975) is another semantic variable which has been of theoretical interest. Concreteness refers to how strongly a word activates perceptual characteristics, and is measured in many studies on a 1 (abstract; language based) to 5 (concrete; experience based) scale (values taken from Brysbaert et al., 2014). Concreteness has been shown to facilitate response times in lexical decision (e.g., James, 1975), which is perhaps not surprising because of its reliance on semantic information, but small effects of concreteness have also been demonstrated in the naming task as well with the conceptually similar variable of imageability (Balota et al., 2004, but see Schock, Cortese, & Khanna, 2012).

1.7 Participant Characteristics

In addition to these three visual word recognition tasks, participants completed several additional measures of interest that are potentially informative with respect to the influence of age on performance. First, they filled out a demographics questionnaire in which they reported their age, education, socio-economic status (objective and subjective), profession, computer proficiency, hours spent on a computer per week, hours spent reading per week, what type of format they read (electronic, print, or both), and whether their reading habits have changed across their lifespan (see Appendix C). This was followed by the Short Blessed Test (Katzman et al., 1983), a 6-item dementia screening test in which errors on each question are counted (with several questions allowing multiple errors, e.g., "Please remember the following: 'John Smith, 1400 Market St., Chicago, Illinois.'", which may count as 5 separate errors). Scores of 0 to 4 are considered normal, 5-9 indicate questionable cognitive impairment, and 10 or more consistent with a dementing disorder.

The participants were then given three vision tests, in which they were presented with a 10-foot or computer distance (24.8 inches, high or low contrast) Snellen card and asked to start with the smallest line of letters they could see and read successively smaller lines of letters until they could no longer read the whole line accurately. For each test (10 foot, computer distance high and low contrast), the dependent variable was the smallest line on which the participant was fully accurate. Furthermore, the high- and low-contrast computer distance tests also had a partial-scoring option, for which participants received credit for letters identified correctly even on lines not fully accurate. For analysis purposes, a factor analysis was computed based on the partial scores for the high- and low-contrast computer distance and an overall score for the 10-foot vision tests, which yielded a vision factor score for each participant. After the vision test, participants completed the Shipley vocabulary test (Shipley, 1946), which consisted of 40 trials of selecting a synonym of the target word from a set of 4 presented alternatives. The dependent variable for analyses was the number of correct trials.

Finally, participants completed the operation span (OSpan; based on Engle, Tuholski, Laughlin, & Conway, 1999) task[2]. Some participants ran out of time and did not complete this task (N = 25), so analyses including this task were conducted on only the 97 participants who did complete it. The OSpan task consisted of simple equation/word pairs for which participants had to indicate whether the equation was correct or incorrect and memorize the word (e.g., $(6 \times 1) - 5 = 2$, BAGEL). Participants saw 54 pairs total, which were grouped into memory sets of 2-6 items (2 sets of 2, 3, and 5 equation-word pairs and 3 sets of 3, 4, and 6 equation-word pairs,

[2] A Stroop color-naming task was also given at the end of the study, but because of time constraints even fewer participants were able to complete it (N = 63). Because of the paucity of data from this task, it will not be discussed further.

randomly intermixed). The measure in this task was total number of to-be-remembered items recalled correctly.

Chapter 2: Method

2.1 Participants

A total of 127 participants participated in the study. Participants completed one of three counterbalance lists (N = 46 in list 1, N = 35 in list 2, and N = 46 in list 3). All participants were recruited through the Washington University Research Participant Registry[3] and paid $10 an hour for their approximately two hours of participation. Participants ranged in age from 18 to 86 (M = 48.7). A histogram of the age distribution is presented in Figure 1, and demographic characteristics are presented in Table 1. Five participants were eliminated for exhibiting outlier characteristics, i.e. difficulty following task instructions (N = 3), non-native English speaker (N = 1), or less than 80% accuracy on naming (N = 1). All other participants were able to follow task instructions, were native English speakers, and achieved greater than 70% accuracy on animacy and lexical decision judgments and greater than 80% accuracy on naming.

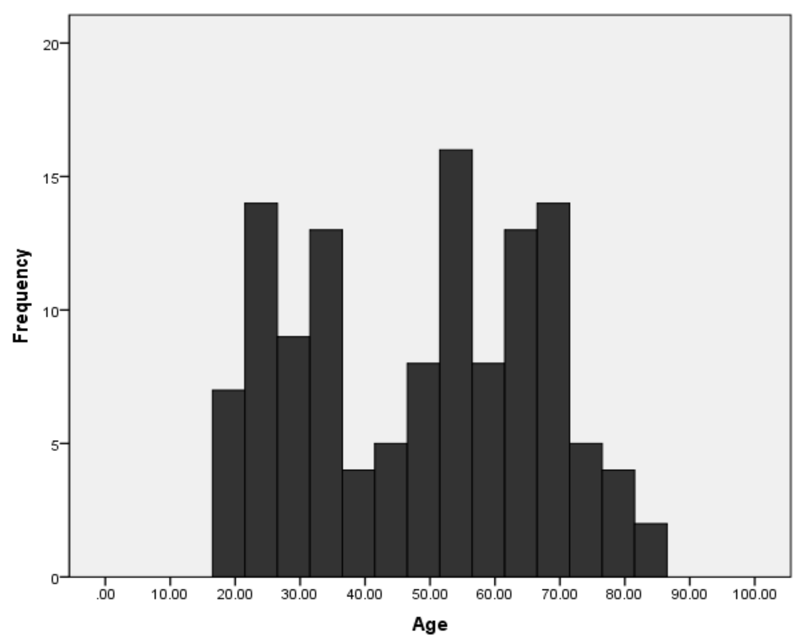

Figure 1. Histogram of Participant Age

[3]The Washington University Research Participant Registry is designed to include a sample of potential participants representative of the general St. Louis area population in terms of race, ethnicity, and socio-economic status, and includes a wide range of ages.

Table 1.
Overall Subject Descriptives (top) and Descriptives split by Age Group (bottom)

Predictor	Mean	Standard Deviation	Minimum	Maximum
Age	48.70	18.29	19.00	86.00
Short Blessed Score	0.81	1.38	0.00	8.00
Vision factor	0.00	1.00	-2.32	1.92
Education	15.52	2.55	9.00	22.00
OSpan	31.01	6.89	15.00	45.00
Shipley Vocab.	32.75	4.12	20.00	39.00
Reading/week	15.75	15.56	0.00	84.00

Predictor	Younger Adults	Middle-Aged Adults	Older Adults
Age	24.62	45.86	68.79
Short Blessed Score	0.48	1.00	0.81
Vision factor	0.87	0.14	-0.79
Education	15.67	15.15	15.87
OSpan	33.65	31.14	28.79
Shipley Vocab.	31.10	31.74	35.10
Reading/week	12.91	18.52	14.44

2.2 Stimuli

A total of 1200 words were selected for use in this study. These stimuli were taken from multiple sources. First, 500 words were taken from Andrews and Heathcote (2001), which were nouns divided equally into non-living and living and high and low frequency (according to Kučera & Francis, 1967). Words that are unknown or inappropriate for American participants (N = 3; *rosella, rostrum, negro*), ambiguous as to animacy (N = 1; *cult*, originally included as an animate noun), not in the ELP (N= 1; *layman*), or repeated in their original stimulus list (N = 2; *visitor* and *gutter* listed twice) were removed, leaving 493 words. Some words were altered to have more traditional American spellings (N = 2; letter “u” removed from *harbour* and *odour*). An additional 975 words were randomly selected from the mono-morphemic nouns (N = 4842) in the SUBTLEX database (Brysbaert & New, 2009) after removing stimuli that would be unknown to American participants (*ninny, cad*) and stimuli of vague animacy (e.g., *almond, parasite*).

In order to further ensure that participants would not have any difficulty making animacy judgments, and to obtain the concordance estimates, a norming study was conducted on Amazon's Mechanical Turk (mturk.amazon.com). One hundred and fifty-seven participants rated animacy of words as "definitely non-living", "mostly non-living", "ambiguous", "mostly living", "definitely living", or "do not know". Participants rated an average of 220 words each, although participants who did not complete the task but rated at least 100 words were included. All participants were from the United States and reported fluency in English. These participants were 49.4% female and had a mean age of 37.8 years (SD = 11.5, range 20-68) and reported a range of highest education level categories (N = 1 for some high school, N = 25 for high school graduate, N = 43 for some college, no degree, N = 16 for associates degree, N = 62 for bachelor's degree, and N = 10 for graduate degree). The age and education ranges help to ensure that these ratings are applicable for participants tested in the current word recognition experiments. This norming procedure resulted in 1468 words, each rated by at least 15 participants. To select the 1200 words for the current study, words were eliminated if over 20% of the participants rated them as ambiguous (N = 53 words) or over 20% of the participants rated them as "do not know" (N = 51 words). Of the remaining words, the 600 living and 600 nonliving words with the highest concordance scores were selected. As noted earlier, concordance was the number of people who rated the word as "definitely" or "mostly" living or non-living divided by the total number of ratings for that word. The animate and inanimate words differed on overall concordance score, $t(1198) = 6.95$, $p < .001$, but the ranges (.67-1.00 for animate; .79-1.00 for inanimate), means (.90 for animate; .94 for inanimate), and standard errors were similar (.004 for animate; .003 for inanimate). For characteristics of the full stimulus set, see Table 2, and for correlations among the item predictor variables and task performance, see Table 3.

Table 2.
Overall Item Predictor Variable Descriptives

Item Predictor Variable	Mean	Standard Deviation	Minimum	Maximum
Length in Letters	6.05	1.74	2.00	13.00
Log SUBTLEX Word Frequency	2.31	0.77	0.30	4.97
Number of Syllables	1.90	0.80	1	5
Number of Morphemes	1.14	1.14	1	3
Orthographic N	3.61	5.48	0.00	34.00
Phonological N	8.70	12.08	0.00	60.00
Orthographic Levenshtein Distance	2.29	0.81	1	6.75
Phonological Levenshtein Distance	2.12	0.92	1	6.65
Concordance	0.92	0.08	0.67	1.00
Consistency	0.66	0.44	0.00	4.96
Valence	5.24	1.18	1.63	8.05
Concreteness	4.23	0.77	1.52	5.00

Table 3

Correlation Matrix for Item Predictor Variables and Overall Task Performance

	1	2	3	4	5	6	7
1. Length	1.00	-.317***	-.019	-.215***	.000	-.084**	.200***
2. Word Frequency		1.00	.201***	-.031	.218***	.036	-.509***
3. Concordance			1	-.113***	.112***	.271***	-.439***
4. Consistency				1	-.046	-.062*	.129***
5. Valence					1	.147***	-.208***
6. Concreteness						1	-.204***
7. AJT RT							1
8. LDT RT							
9. NMG RT							
10. AJT ACC							

	8	9	10	11	12
1. Length	.456***	.562***	-.035	-.077**	-.200***
2. Word Frequency	-.667***	-.585***	.278***	.492***	.311***
3. Concordance	-.205***	-.163***	.581***	.271***	.171***
4. FF Rime Consistency	.129***	.027	-.074*	-.073*	-.224***
5. Valence	-.208***	-.207***	-.147***	.085**	.138***
6. Concreteness	-.204***	-.196***	-.173***	-.009	.113***
7. AJT RT	1	.570***	.522***	-.539***	-.476***
8. LDT RT		1	.755***	-.234***	-.594***
9. NMG RT			1	-.236***	-.463***
10. AJT ACC				1	.292***
11. LDT ACC					1
12. NMG ACC					

Note. AJT = Animacy Judgment Task, LDT = Lexical Decision Task, NMG = Naming Task, OLD = Orthographic Levenshtein Distance, PLD = Orthographic Levenshtein Distance. + p < .10. * p < .05. ** p < .01. *** p < .001.

Three counterbalancing lists were created using a random number generator to rotate through the three different tasks, with the caveat that each list required 200 animate and 200 inanimate words. A one-way ANOVA indicated that the lists did not differ significantly on concordance, length, raw or LOG HAL or SUBTLEX word frequency, orthographic or phonological N, orthographic or phonological Levenshtein distance, # of syllables, or # of morphemes (*p*s > .05)[4].

Nonword stimuli (N = 400) for the lexical decision task were generated by the Nonword Generator Wuggy (Keuleers & Brysbaert, 2010), which segmented the 1200 word stimuli into syllables and recombined them to create 400 nonwords. All participants saw the same 400 nonwords, which were equated with the words on length in letters, t (1583) = .52, $p = .603$. As expected, words and nonwords differed on orthographic N, t (1583) = 3.51, $p < .001$. Word and nonword stimuli are listed in Appendices A and B, respectively.

2.3 Procedure

Participants completed three visual word recognition tasks and several other tasks in a single two-hour experimental session in the following order: Short Blessed Test as a cognitive screening measure, animacy judgment, vision test, word naming, Shipley vocabulary, lexical decision, and OSpan. Order was held constant because emphasis here is on individual differences and so this minimizes variability due to counterbalancing order. Within each word recognition task participants completed 12 practice trials in each task, followed by 400 trials in the animacy

[4] Although the counterbalance lists were selected to be equal on concordance, length, raw and log HAL frequency, and raw SUBTLEX word frequency, orthographic and phonological N, orthographic and phonological Levenshtein distance (OLD/PLD), number of syllables, and number of morphemes, further analyses indicated that the lists showed some differences with respect to item- and subject-level performance. In the item-level analyses, PLD, valence, and consistency interacted with list in animacy judgment, and length and word frequency interacted with list in lexical decision. In the subject-level analyses, there was an effect of list for length, PLD, and consistency in animacy judgment, length, word frequency, OLD, concordance, consistency, and concreteness in lexical decision, and length, concordance, valence, and concreteness for naming. These list effects highlight the influence of stimulus selection and list context on different visual word recognition tasks. For the purposes of the current study, all lists will be collapsed across to mitigate these individual list effects and increase power.

judgment and naming tasks and 800 trials in the lexical decision task. Stimuli were rotated through the tasks within three counterbalanced lists so that each participant saw each word only once across all tasks, and words were presented in a random order.

On each trial in each of the visual word recognition tasks, participants first saw a 400-millisecond (ms) fixation cross at the center of the screen to indicate that the trial was about to begin. The stimulus then appeared and participants were instructed to provide the appropriate response (reading the word aloud for the word naming task, pressing a key corresponding to a "word" or "nonword" decision for the lexical decision task, or pressing a key corresponding to a "living" or "nonliving" decision for the animacy judgment task). The stimulus remained on the screen until a vocal (microphone) or key press response was detected, at which point a 200-ms blank screen appeared until the start of the next trial.

Chapter 3: Results

In order to ensure that extreme scores did not strongly influence the results, the following outlier procedures were used. First, for the animacy judgment task (which produced relatively slower RTs, see Table 4 for overall task performance), trials that produced response latencies below 250 ms or above 4000 milliseconds (ms) were removed (0.66% of trials). Trials below 250 ms and over 3000 ms were removed for the lexical decision task (1.1% of correct trials) and the naming task (0.17% of correct trials). Microphone errors (invalid triggering of the microphone on a trial, e.g., coughing or stammering) were also removed for the naming task. RTs were then converted to *z*-scores, which transforms response latency on each trial onto a standardized scale based on the mean and standard deviation of that individual participant. This accounts for the well-documented general slowing that occurs across the lifespan (Faust, Balota, Spieler, & Ferraro, 1999). Trials outside of three standard deviations (SDs) from the mean were then removed from the remaining trials, which resulted in removing a further 2.2% of trials for animacy judgment, 2.3% of trials for lexical decision, and 1.6% of trials for naming. Furthermore, 15 words were eliminated because mean performance across participants was less than 50% in one of the tasks (animacy judgment: hemlock, petal, limb, cell, grape, bush, barnacle, lark, and thigh; lexical decision: frigate; naming: brasserie, anemone, soot, dachshund, cellist). The total percentage of trials included in the following RT analyses was 91% of all observations for animacy judgment, 92% for lexical decision, and 93% for naming.

Table 4.

Overall task performance

	Mean Raw RT	RT Standard Deviation	Mean Accuracy	Accuracy Standard Deviation
Animacy Judgment	955	124	.94	.09
Lexical Decision	791	97	.96	.07
Naming	570	55	.98	.06

3.1 Comparison of Different Word Frequency Metrics

Four different word frequency metrics were examined for potential use in foregoing analyses: Kučera and Francis (K&F; Kučera & Francis, 1967), Hyperspace Analogue to Language (HAL; Lund & Burgess, 1996), the Educator's Word Frequency Guide (Zeno; Zeno, Ivens, Millard, & Duvvuri, 1995), and Subtitle Frequency (SUBTLEX; Brysbaert & New, 2009). All metrics were transformed to log functions because such transformations have been shown to better fit the response latencies (see for example Adelman, Brown, & Quesada, 2006). Because word frequency metrics are influenced by the corpora from which they are drawn, the metrics might predict differential amounts of variance across the age groups. For example, K&F is derived from a small corpus, about 1 million words. HAL, SUBTLEX, and Zeno were drawn from much larger corpora than K&F, so they should be more suitable word frequency norms in general (Brysbaert & New, 2009). However, because age is considered in the current study, it must also be considered with respect to the word frequency norms. Specifically, K&F was assembled from texts dating from 1961, HAL from internet posts, SUBTLEX from movie subtitles, and Zeno from young adult texts (grades 1-12) taken prior to 1995. Therefore, as noted earlier, it is possible that the relative predictive power for these metrics differs across the age span, such that HAL and Zeno are most predictive for young adults, whereas SUBTLEX captures equal variance for all participants. One might even predict that K&F would well predict older adults performance (although Balota et al., 2004, found that it was a poor measure for younger and older adults).

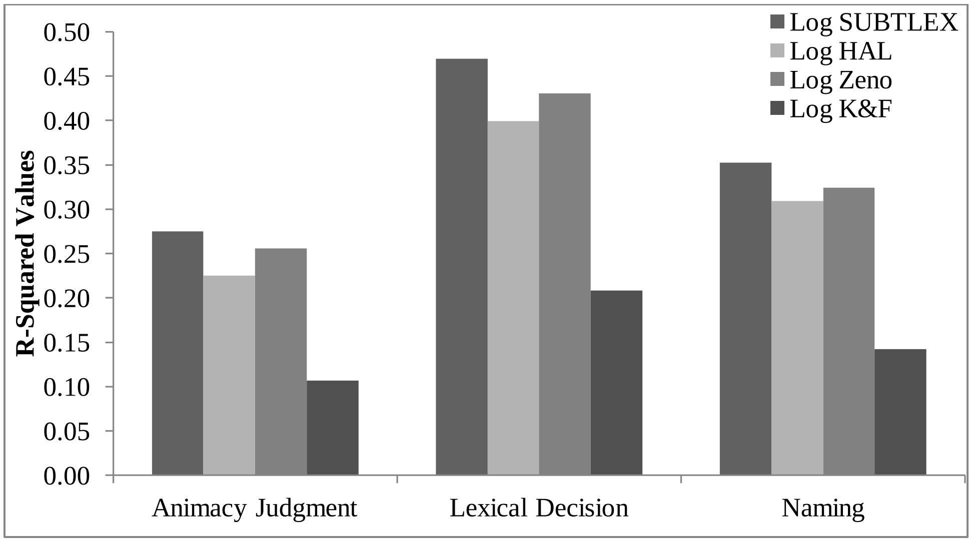

Figure 2. R^2 values for each word frequency metric when entered alone in model (as in Balota et al., 2004).

To assess the relative predictive power of the word frequency metrics, each metric was entered as the sole predictor variable in a regression analysis predicting *z*-scores for each task separately, as in Balota et al. (2004) (see Fig. 2). SUBTLEX outperformed the other metrics across all tasks, although differences among SUBTLEX, HAL, and Zeno were small. K&F had drastically less predictive power, at less than half the R^2 as the other metrics. Next, each word frequency metric was included, one at a time, in the overall item-level regression analyses (see Fig. 3). In this analysis, SUBTLEX and Zeno performed similarly to one another, with HAL slightly less robust in the animacy judgment and lexical decision, and K&F as the least predictive metric across all three tasks.

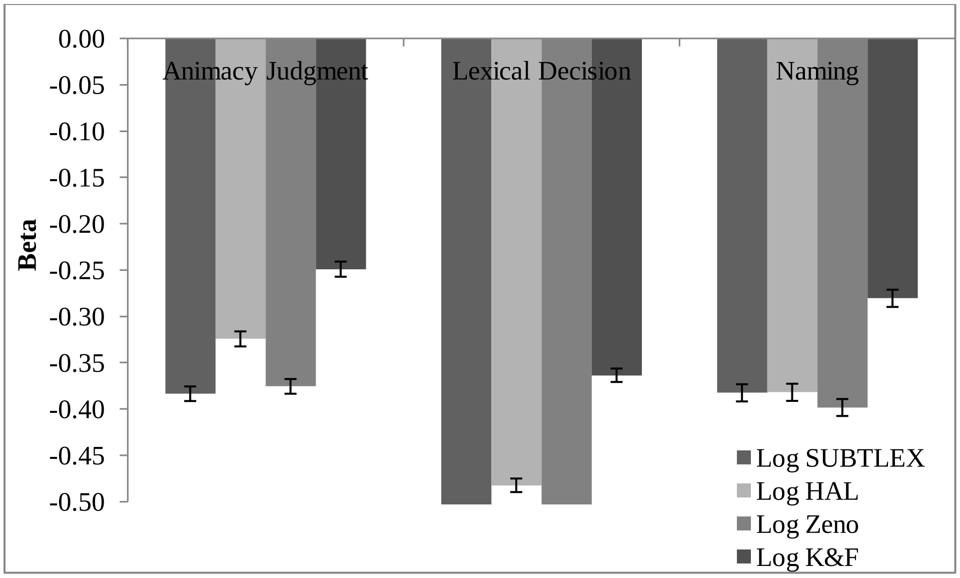

Figure 3. Beta weights for each word frequency metric when entered (one at a time) in model with other predictor variables from full item-level regression.

Figure 4 displays the beta weights for each of three age groups (here, younger adults included ages 18 – 30, middle-aged adults included ages 31 – 59, and older adults included ages 60+) as a function of task. As shown, SUBTLEX and Zeno were the most predictive measures of word frequency, HAL slightly less predictive, and K&F least predictive. One interesting observation is that HAL was as predictive as SUBTLEX and Zeno for young adults only, whereas it was inferior to SUBTLEX and Zeno for middle-aged and older adults. This is potentially because HAL is based on internet posts, which may be more reflective of the young adult lexicon; however, one might expect to see this pattern for Zeno as well, which is based on young adult texts. Because log SUBTLEX performed relatively well, and Brysbaert and New (2009) have argued it accounts for the most variance in large scale databases, this measure was used in all further analyses.

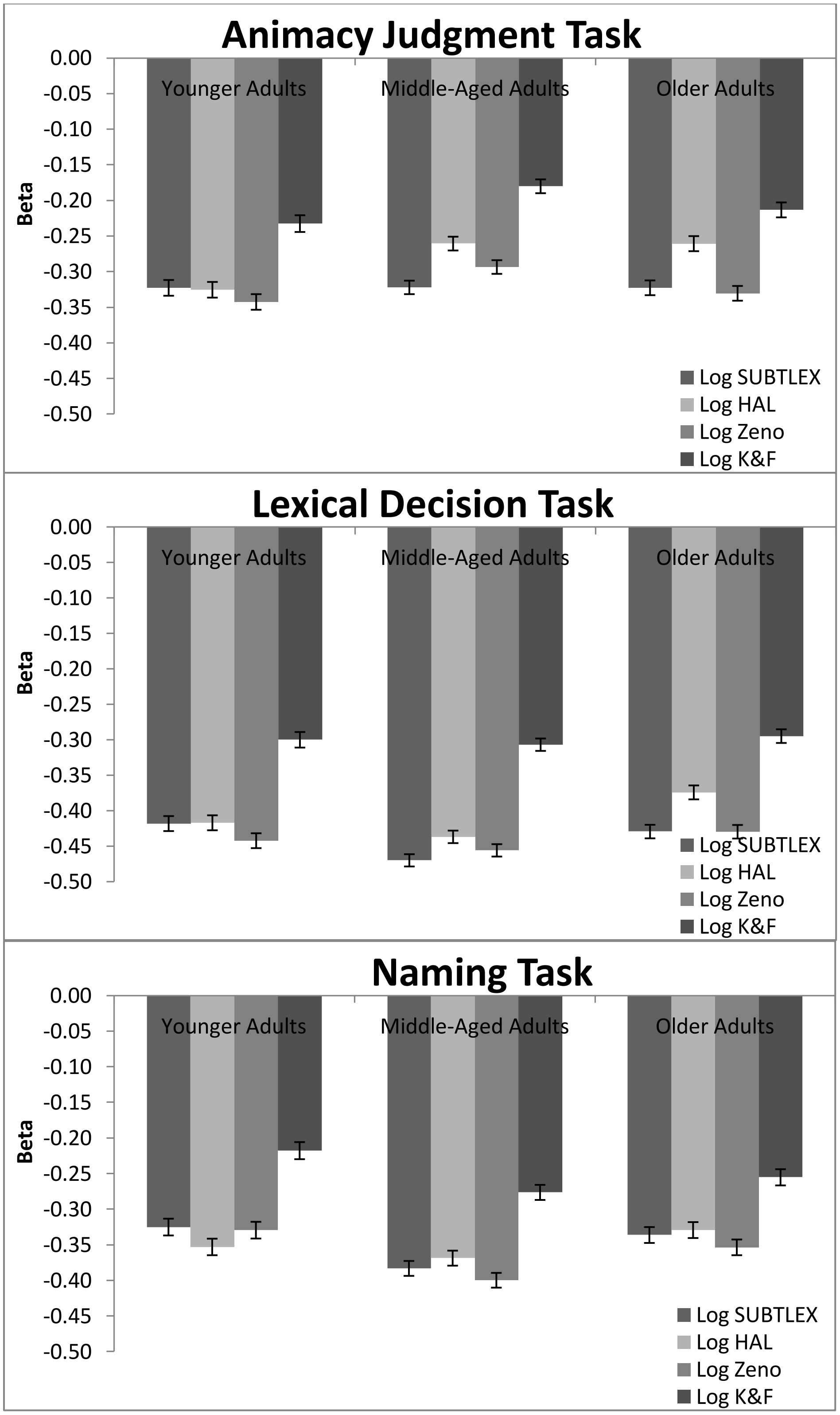

Figure 4. Word frequency metrics separately for each task and age group when entered (one at a time) in model with other predictor variables from full item-level regression.

3.2 Data Analysis: General Approach

Three main approaches to data analysis are reported: item-level, subject-level, and mixed effects modeling. The item-level analyses, using items as the basic units of analysis, allow examination of the general effects of the predictor variables on performance in the three tasks. These are important initial analyses to make contact with the extant literature to insure that one finds the standard pattern of effects in the variables measured in the current study. The subject-level analyses involved conducting an analysis on each subject to obtain beta weights for each predictor variable in each task, which allows for consideration of the critical questions regarding how age and correlated variables may influence the word frequency effect. Finally, the linear mixed effects modeling technique includes item and participant random slopes and models trial-level data with item- and subject-level predictors, enabling a consideration of age and word frequency in the same regression model. *Z*-scores and accuracy (proportion correct out of total trials, not including microphone errors in the naming task) are reported here, with the exception of the linear mixed effects analyses which used raw RT and accuracy. Analyses of the raw RTs in the item- and subject-level analyses produced the same overall patterns as *z*-scores in most analyses with the exception of a main effect of age which is not present in *z*-scores because of the *z*-score correction for age-related slowing. Participants approached ceiling performance on accuracy (94% for animacy judgment, 96% for lexical decision, and 98% for naming), making accuracy measures highly skewed and thus of questionable interpretive value.

The first step of the item- and subject-level regression analyses contained a set of 13 predictors to represent phonological onsets. Step two contained length in letters, log SUBTLEX word frequency (Brysbaert & New, 2009), concordance (from the Mturk ratings), consistency taken from Yap and Balota (2009), valence (Warriner et al., 2013), and concreteness (Brysbaert et al., 2014).

3.3 Item-Level Analyses

Phonological onsets. The first step of the item-level regression included phonological onsets (see Tables 5 and 6; for brevity, beta weights for the individual onset characteristics are not presented). This step should primarily have an influence on the naming task, but it predicted significant variance in the lexical decision task *z*-scores as well, *p*s < .05. However, phonological onsets did predict *more* variance in the naming task (R^2 = .049 for *z*-scores, R^2 = .062 for accuracy) than the lexical decision task (R^2 = .024 for *z*-scores, R^2 = .041 for accuracy). The 4.9% of variance predicted by phonological onset variables in the naming task is comparable to the 4.3% found by Yap and Balota (2009). However, this is substantially less than the ~35% variance which has been reported on large sets of monosyllabic words (e.g., Balota et al., 2007; Yap & Balota, 2009).

Table 5.
Item-Level Results for Z-Scores across Tasks

	Animacy Judgment		Lexical Decision		Naming	
Predictor	ΔR^2	Beta	ΔR^2	Beta	ΔR^2	Beta
Step 1 Phonological Onsets	.016		.024*		.049***	
Step 2	.359***		.472***		.471***	
Length		.072*		.271***		.439***
Word Freq		-.384***		-.520***		-.383***
Concordance		-.339***		-.035		-.050*
Consistency		.066*		.019		-.032
Valence		-.080**		-.083***		-.018
Concreteness		-.079**		-.144***		-.134***
Total R^2	.375***		.496***		.520***	

Note. + *p* < .10. * *p* < .05. ** *p* < .01. *** *p* < .001.

Table 6.

Item-Level Results for Accuracy across Tasks

	Animacy Judgment		Lexical Decision		Naming	
Predictor	Δ R^2	Beta	Δ R^2	Beta	Δ R^2	Beta
Step 1 Phonological Onsets	.007		.041***		.062***	
Step 2	.381***		.189***		.090***	
Length		.007		.047		-.123***
Word Freq		.171***		.390***		.181***
Concordance		.603***		.126***		.045
Consistency		.006		-.078*		-.129***
Valence		.000		.023		-.036
Concreteness		-.189***		.057+		.142***
Total R^2	.388***		.252***		.158***	

Note. + p < .10. * p < .05. ** p < .01. *** p < .001.

Yap and Balota (2009) suggested that phonological onsets are more influential in monosyllabic words because they explain a set amount of variance and there is less overall variance to explain in monosyllabic words. Hence, onsets explain *proportionally* more for monosyllabic words than for multisyllabic words, which have many other variables contributing to their performance. To explore this in the current dataset, phonological onsets were considered for monosyllabic words (N = 386) and multisyllabic words (N = 687) separately (see Table 7). For lexical decision and naming *z*-scores, onsets explained more variance in monosyllabic words (R^2 = .107, R^2 = .151, respectively, *p*s < .001) than multisyllabic words (R^2 = .036, R^2 = .043, respectively, *p*s < .05). This pattern persisted in accuracy for lexical decision; R^2 = .208 for monosyllabic words and R^2 = .050 for multisyllabic words (*p*s < .01), but the opposite was true for naming accuracy, in which the phonological onset regression step was not significant for monosyllabic words (p = .845) and significant for multisyllabic words (R^2 = .282, p < .001). The phonological onsets were not significant for animacy judgment in z-scores or accuracy (ps > .05) except when analyses only focused on multisyllabic words, which did produce a reliable effect in z-scores, R^2 = .038, p < .05.

These results are broadly consistent with the monosyllabic versus multisyllabic distinction made previously (Yap & Balota, 2009), but it is surprising that phonological onsets did not explain more variance in the naming task; the magnitude of the variance explained in our naming task versus, for example, Yap and Balota (2009) is considerably lower. To explore this further, mono- versus multi-syllabic z-scores from the English Lexicon Project (ELP) for the items used in the present study were subjected to the same item-level regression analyses. This showed similar magnitude effects of phonological onsets in lexical decision (R^2 = .083 for monosyllabic words, R^2 = .034 for multisyllabic words, ps < .05), but much larger effects of phonological onsets in naming (R^2 = .432 for monosyllabic words, R^2 = .112 for multisyllabic words) relative to the current study. It seems that the muted effect of phonological onsets in naming in the current study is not due to the stimuli, since the effect is much larger in the ELP data for the same set of items, but more likely due to some characteristic of the voicekey used in the present study or participant characteristics, since the ELP data is primarily on younger adults.

Table 7.

R^2 Values for All Phonological Onsets Combined, Split by Mono- vs. Multisyllabic

	Z-scores		Accuracy	
	Monosyllabic, R^2	Multisyllabic, R^2	Monosyllabic, R^2	Multisyllabic, R^2
Animacy Judgment	.038	.038*	.027	.011
Lexical Decision	.107***	.036*	.208***	.050**
Naming	.151***	.043*	.021	.282***

Lexical and Semantic Predictor Variables: *Z*-scores. The second step of the item-level regression included length, word frequency, concordance, consistency, valence, and concreteness (see again Table 5, and Fig. 5). This step predicted more variance in *z*-scores than the first step across animacy judgment ($\Delta R^2 = .359$), lexical decision ($\Delta R^2 = .472$), and naming ($\Delta R^2 = .471$). Variance predicted in both steps is of comparable magnitude to similar studies (e.g., Balota et al., 2004; Yap & Balota, 2009). As shown, and consistent with prior literature, the influence of many of the predictor variables depended on the task.

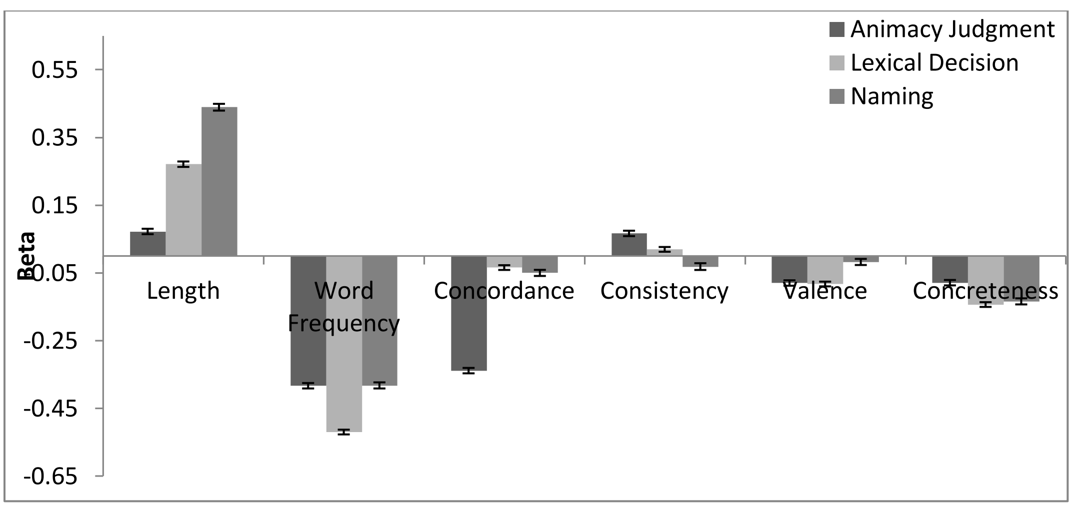

Figure 5. Item-Level Beta Weights for *z*-scores. Error bars denote standard error of the mean.

Word frequency effects were robust across all three tasks, but were largest for lexical decision, with animacy judgment and naming producing comparable, and smaller effects. As discussed in the introduction, large effects of word frequency in lexical decision are consistent with prior literature, because frequency is helpful in the word/nonword decision, i.e., being diagnostic of words but not nonwords. In this light, word frequency may influence both the word identification stage of task performance *and* the decision stage of task performance (e.g., Andrews & Heathcote, 2001; Balota & Chumbley, 1985). Word frequency was also robust in the animacy judgment, which is consistent with some prior literature also using a binary judgment (Andrews & Heathcote, 2001; Forster & Shen, 1996; Monsell, Doyle, & Haggard, 1989; but see Balota & Chumbley, 1984; Forster, 1985).

Length in letters had a significant inhibitory effect on all three tasks, but was most robust in naming, followed by lexical decision, with the smallest effect in animacy judgment. Consistency was slightly inhibitory for animacy judgment only. Naming should show the strongest effects of consistency on performance, on the basis of the orthography-to-phonology computation it requires, so this effect was puzzling. The consistency measure is discussed further in the General Discussion.

Turning to the semantic variables, concordance was facilitatory, significant, and robust in the animacy judgment task; in fact, it was nearly as strong a predictor as word frequency. This was predicted because of the measure's direct relevance to the animacy judgment task. It is interesting that concordance produced a small reliable effect in naming also. Valence (coded so that higher values are positive) showed small facilitatory effects in the animacy judgment and lexical decision tasks, but not in the naming task. This is reflective of a positivity bias, and is consistent with the idea that negatively valenced words induce automatic vigilance and difficulty disengaging from them (Algom et al., 2004; Estes & Adelman, 2008b), but it does reflect the greater degree of semantic activation in the animacy judgment and lexical decision tasks relative to naming. Concreteness produced robust facilitatory effects of similar magnitude in lexical decision and naming but had a much smaller influence in animacy judgment. Regardless of the specific magnitudes of effects across tasks, the influence of concreteness is a clear demonstration of a semantic effect on visual word recognition or decision processes (Brysbaert, Warriner, & Kuperman, 2013; James, 1975; Whaley, 1978), and the effect of emotional valence suggests that lexical processing can also be influenced by emotional content (Augustine, Mehl, & Larsen, 2011; Warriner et al., 2013).

Lexical and Semantic Predictor Variables: Accuracy. For accuracy, as for *z*-scores, the second step of the item regression predicted more variance than the first step across animacy judgment ($\Delta R^2 = .381$), lexical decision ($\Delta R^2 = .189$), and naming ($\Delta R^2 = .090$). The overall proportion of variance explained, and the effects of predictor variables on accuracy performance were considerably smaller than the effects on *z*-scores; however, there were several significant effects of predictor variables on accuracy (see Table 6 and Figure 6). These are also described below to some extent but are interpreted cautiously because of the relatively large skewing of the accuracy data[5].

[5] In the reported analyses, accuracy was calculated as the number of correct trials out of total trials not including outliers and microphone errors. However, a different calculation was also explored to try to minimize the extreme skewing in accuracy: number of correct trials out of total trials presented, so in this case including outliers and microphone errors in the denominator. Although this did reduce overall accuracy (and skewing) somewhat, the distributions were still highly skewed, and in fact correlated with the original measure at r = .87 to r = .96 (and therefore showed the same patterns as the previous accuracy measure). A last attempt on the item-level data was to log transform the first accuracy measure beta weights, but this did not substantially alter the patterns either.

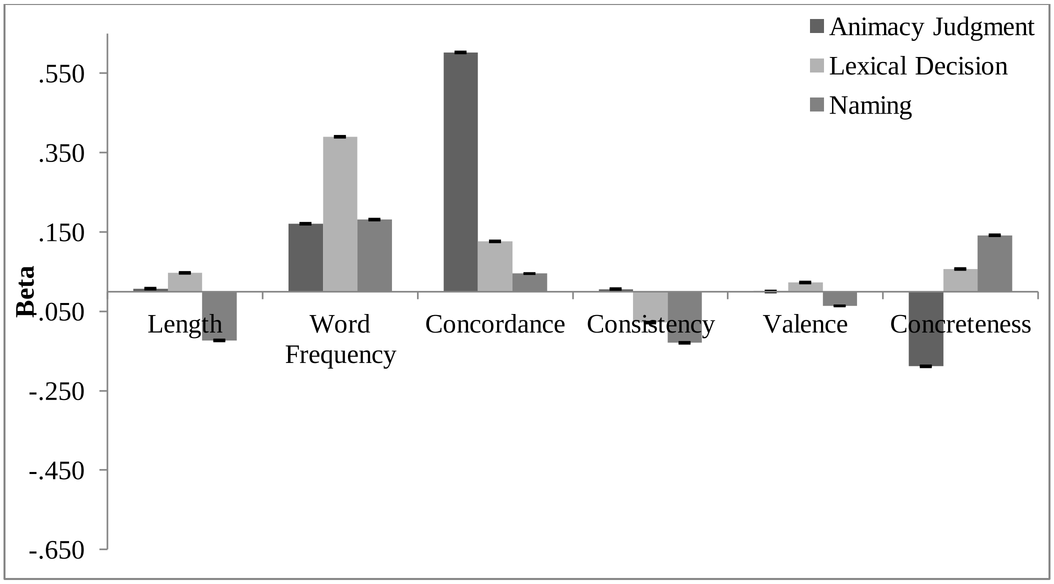

Figure 6. Item-Level Beta Weights for accuracy. Error bars denote standard error of the mean.

Word frequency was a robust predictor across all three tasks. Like the *z*-score analyses, lexical decision shows the largest effect of word frequency with animacy and naming showing similar, more modest effects of word frequency.

Length was significant and inhibitory for naming only, which accords well with the finding of the strongest effect on *z*-scores in naming. Consistency was significantly inhibitory for lexical decision, and even more strongly inhibitory for naming.

Turning to the semantic variables, concordance was significant and facilitatory for animacy judgment; in fact, it was the most robust predictor, greater than even word frequency in this case. It was also significant, albeit much more modest, for lexical decision. Valence was not significant for any task, but concreteness was significant and inhibitory for animacy judgment, but facilitatory for lexical decision, albeit marginally significant, and naming.

3.4 Subject-Level Analyses

Subject-level predictors are displayed in a correlation table below (Table 8).

Table 8.
Correlation Matrix for Subject Predictor Variables

	Age	Age Quadratic	Short Blessed Test	Vision Factor	Education	Ospan	Shipley	Reading per week
Age		.987***	.032	.682***	.071	-.297**	.400***	.029
Age Quadratic		1	.011	.677***	.077	-.276**	.414***	.015
Short Blessed Test			1	.153+	-.201*	-.332**	-.225*	.023
Vision- Average				1	-.097	-.170	.156+	-.099
Education					1	.234*	.447***	.092
Ospan						1	.338**	.043
Shipley							1	.115
Reading per week								1
AJT RT								
LDT RT								
NMG RT								
AJT ACC								
LDT ACC								
NMG ACC								

	AJT RT	LDT RT	NMG RT	AJT ACC	LDT ACC	NMG ACC
Age	.252**	.301**	.105	.369***	.366***	.125
Age Quadratic	.206*	.268**	.082	.381***	.382***	.160+
Short Blessed Test	.173+	.138	.336***	-.233*	-.272**	-.326***
Vision- Average	.248**	.289**	.228*	.085	.031	-.139
Education	-.123	-.104	-.310***	.404***	.407***	.373***
Ospan	-.295**	-.246*	-.375***	.299**	.212*	.212*
Shipley	-.192*	-.145	-.376***	.717***	.666***	.600***
Reading per week	-.021	-.085	-.145	.251**	.230*	.133
AJT RT	1	.777***	.399***	-.123	-.161+	-.205*
LDT RT		1	.563***	-.031	-.170+	-.182*
NMG RT			1	-.361***	-.515***	-.404***
AJT ACC				1	.743***	.592***
LDT ACC					1	.619***
NMG ACC						1

Note. AJT = Animacy Judgment Task, LDT = Lexical Decision Task, NMG = Naming Task, + $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

***Z*-scores.** As noted, item-level regressions were conducted on each participant to obtain beta weights for each participant. This allows examination of correlations between word frequency betas for *z*-scores and accuracy and the critical participant characteristics of interest (e.g., age, vocabulary, vision, etc.). First, simple bivariate correlations between age and the subject-level *z*-score word frequency betas were all non-significant (*p*s > .49, see Table 9). This indicates that there was no linear relationship between age and the word frequency effect in any task. Because age was considered continuously in this study, examining the quadratic age functions are also important. Specifically, it is possible that the word-frequency effect does not change until one reaches the advanced ages. Correlations between quadratic age and word frequency betas with linear age partialled out also were not significant, except for a marginally significant correlation with word frequency in the naming task ($r = .176$, $p = .053$). This relationship reflected slightly smaller word frequency effects on the extreme ends of the age spectrum, with middle-aged adults showing larger word frequency effects (see Fig. 7).

Table 9

Subject-Level Correlations with Word Frequency Betas

Subject-level Predictors	*Z*-score Betas			Accuracy Betas		
	AJT WF	LDT WF	NMG WF	AJT WF	LDT WF	NMG WF
Age	.015	-.032	.063	.031	-.028	.223*
Age Quadratic	.013	-.029	.091	.016	-.013	.241*
Short Blessed Test Score	-.175+	-.084	-.260**	-.041	-.394***	-.034
Vision	.062	-.081	.084	.015	.094	-.191*
Education	.354***	.174+	.235*	.080	.123	-.101
Ospan	.182+	.248*	.234*	-.081	.094	-.095
Shipley Vocabulary	.362***	.130	.413***	.052	.078	.090
Reading per week	.117	.126	-.059	-.134	-.110	-.017

Note. AJT = Animacy Judgment Task, LDT = Lexical Decision Task, NMG = Naming Task, + $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

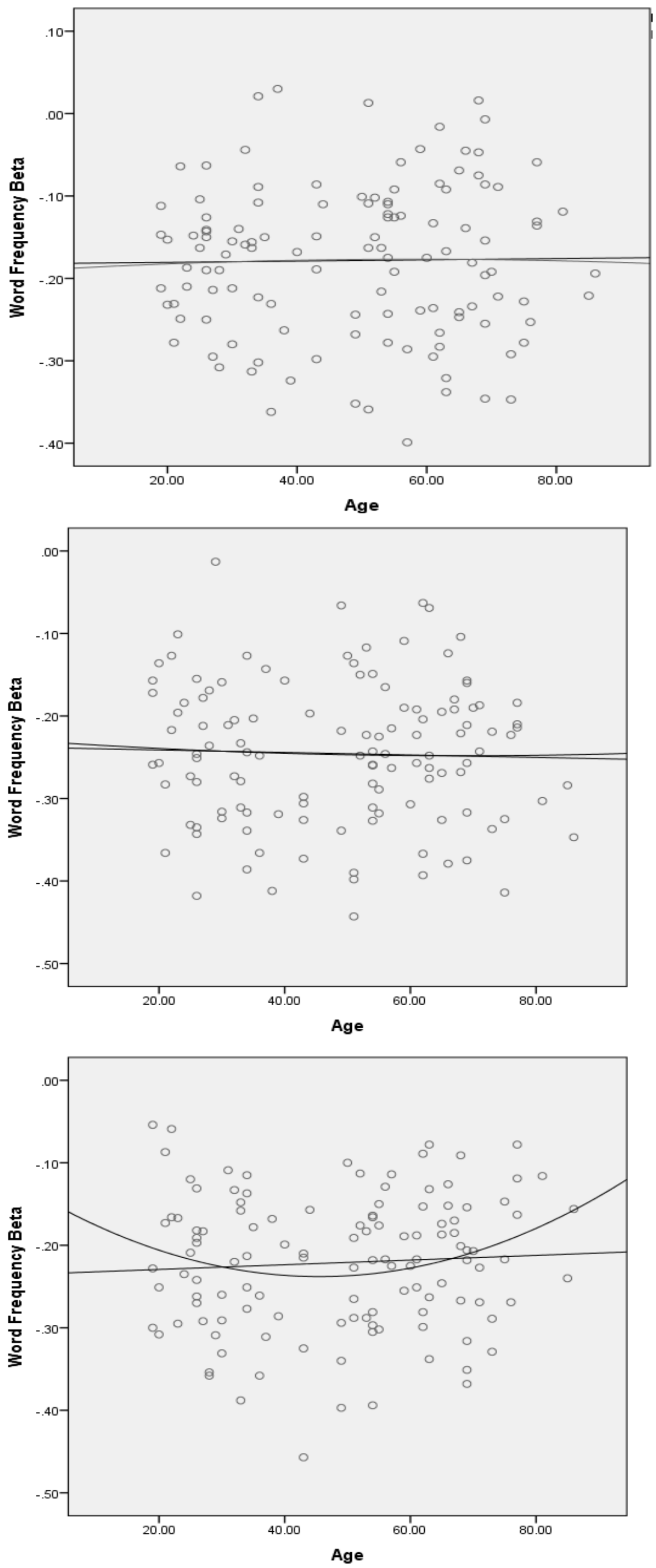

Figure 7. Linear and Quadratic Effects of Age on Word Frequency *Z*-score Betas, Animacy Judgment Task (top), Lexical Decision Task (middle), and Naming Task (bottom)

Of course, a critical issue addressed in the present dissertation is the extent to which demographic characteristics may modulate the relationship between age and word frequency. As noted in the introduction, differences across previous studies as to which demographic characteristics were measured and controlled for between younger and older adults (e.g., vocabulary) may explain the disparate findings in prior literature. Of course, if these variables are not related to age, then it would not be critical to control for them. The intercorrelations among the demographic variables and age are displayed in Table 8. Vision and OSpan were negatively correlated with age, whereas vocabulary was positively correlated with age. These findings are consistent with prior literature (see Schieber, 2006, for age differences in vision, and Hasher & Zacks, 1988; Salthouse & Babcock, 1991, for age differences in working memory), except that age and education were not correlated in the current sample are typically positively associated in prior literature (as in a meta-analysis by Verhaeghen, 2003).

Importantly, controlling for each of these variables separately did not change the observed lack of correlation between age and word frequency beta weights in the three tasks, except for vision linear and quadratic. Specifically, controlling for vision led to a significant correlation between age and word frequency in the naming task (r = .200, $p = .030$, see Table 10 and Fig. 8). Because word frequency is primarily a negative effect, this actually reflects *smaller* word frequency effects with increasing age. Additionally, controlling for Shipley vocabulary score revealed a marginally significant correlation between age and word frequency effects in the animacy judgment task, r = -.155, $p = .091$, as did controlling for the quadratic effect of Shipley vocabulary score, r = -.178, $p = .053$. This negative correlation reflects stronger word frequency effects with increasing age, and it may be important to note that this pattern showed up numerically in the other tasks but did not reach significance (see Table 10).

Table 10.

Subject-Level Partial Correlations between Age and Word Frequency Betas

	Correlations Between Age and Word Frequency					
	Z-Score Beta Weights			Accuracy B Weights		
Control Variables	AJT WF	LDT WF	NMG WF	AJT WF	LDT WF	NMG WF
None (Age only)	.015	-.032	.063	.031	-.028	.223*
Short Blessed Test	.023	-.032	.081	.196*	.031	-.006
Short Blessed Test Quadratic	.025	-.033	.082	.195*	.030	-.005
Vision	.089	-.076	.200*	.097	.064	.045
Vision Quadratic	.077	-.101	.182*	.201*	.035	-.034
Education	-.035	-.037	.025	.189+	.025	-.037
Education Quadratic	-.042	-.039	.047	.191+	.024	-.038
Ospan	.018	.033	.165	.179	-.088	-.010
Ospan Quadratic	.016	.029	.167	.178	-.081	-.016
Vocabulary	-.155+	-.093	-.124	.192*	.009	-.063
Vocabulary Quadratic	-.178+	-.102	-.131	.194*	.008	-.066
Reading per week	.009	-.013	.055	.228*	-.096	-.012
Reading per week Quadratic	.017	-.023	.062	.226*	-.101	-.020
Vision, Education, Vocabulary	-.090	-.136	.007	.075	.046	.016
Vision, Ospan, Vocabulary	-.075	-.091	.086	.042	-.072	.040

Note. AJT WF= Animacy Judgment Task Word Frequency Betas, LDT WF = Lexical Decision Task Word Frequency Betas, NMG WF = Naming Task Word Frequency Betas, + $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

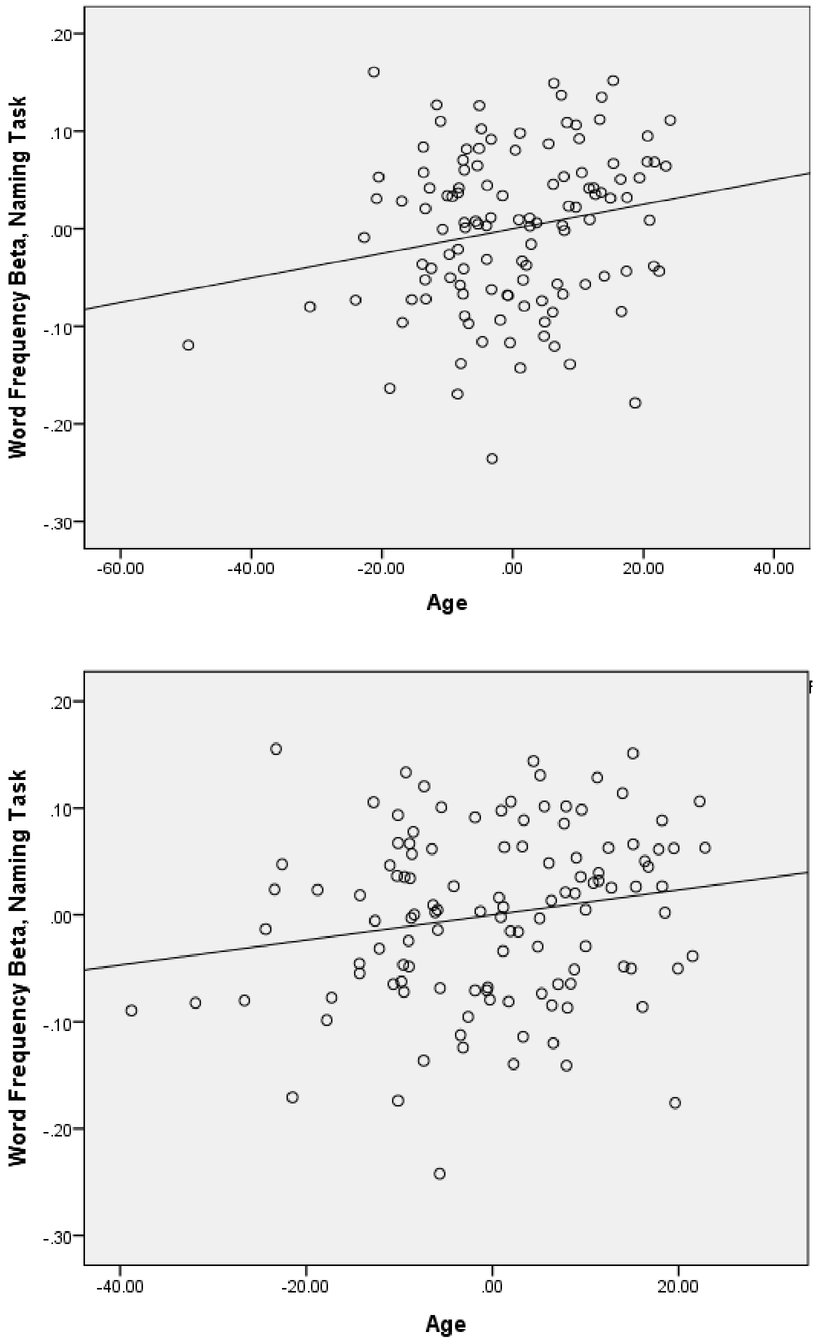

Figure 8. Relationship between Age and Word Frequency *Z*-score Betas in the Naming Task, Controlling for Vision Linear (top) and Vision Quadratic (bottom)

Furthermore, controlling for combinations of demographic variables, including vision, education, and Shipley, or vision, OSpan, and Shipley, did not reveal any correlations between age and word frequency betas.

Table 11.
Subject-Level Correlations with Other Predictor Variable Betas, Z-scores

Correlations: *Z*-scores	Value
Age & AJT Length	-.314**
Vision & AJT Length	-.248**
Shipley & AJT Length	-.359***
Reading Per Week & AJT Valence	-.243**

Correlations: Accuracy	Value
Age & NMG Length	.244**
Short Blessed Test & LDT Length	-.333***
Short Blessed Test & LDT Concordance	-.480***
Short Blessed Test & LDT Concreteness	.321***
Short Blessed Test & LDT Consistency	.458***

Note. Only correlations significant at $p < .01$ level are displayed. AJT = animacy judgment task, LDT = lexical decision task, and NMG = naming task.

An additional issue that one can address in the current study is the extent to which other item predictor variables change as a function of age and other demographic variables. For example, it is possible that the effect of length is modulated by age and vision. Table 11 displays correlations between demographic characteristics and subject betas for other predictor variables. Because this analysis is exploratory and involves many comparisons, reported below are only those which reached significance at the $p < .01$ level.

First, consider the correlations in Table 9. Word frequency in the naming task is correlated with Short Blessed Test score, r = -.260, $p = .004$, such that higher general cognition is associated with smaller word frequency effects. Word frequency in the animacy judgment task is correlated with education, r = .354, $p < .001$, and Shipley vocabulary, r = .362, $p < .001$, and word frequency in the naming task is correlated with Shipley vocabulary, r = .413, $p < .001$[6]. All of these correlations reflect relatively smaller word frequency effects, even approaching null, as the participant characteristic, education or vocabulary, gets higher. Next, as shown in the top half of Table 11, age was not associated with changes in any variables except for a smaller length effect in the animacy judgment task, r = -.314, $p < .001$, in the *z*-score betas. The negative relation between age and length effects is consistent with Spieler and Balota's (2000) finding of smaller length effect in the older than the younger adults with the naming task, but it is unclear why this relationship would emerge in only the animacy judgment task. Furthermore, vision was associated with length in the lexical decision task, r = -.248, $p = .007$, with length betas approaching 0 at higher vision scores. Shipley vocabulary was associated with length in the animacy judgment task, r = -.359, $p < .001$, again with length betas approaching zero at higher vocabulary scores. This type of effect, in concert with the findings for word frequency betas with vocabulary and education described above, may be considered hallmarks of the skilled lexical processor (i.e., those with higher vocabulary process words more automatically and thus are less influenced by word variables, e.g., LaBerge & Samuels, 1974; Stanovich, 1980). One perhaps might have expected effects to show up more broadly, e.g., vocabulary should be associated with declines in most lexical variables, not just length and word frequency, but these variables are the most robust predictors. Finally, hours read per week was associated with valence in the animacy judgment task, r = -.243, $p = .009$. There is a lot of variability in valence effects, with some

[6] The correlation between Shipley vocabulary and word frequency in the lexical decision task did not reach significance, r = .130, $p = 156$, but this seemed to be driven by an outlier participant who did not show a word frequency effect in lexical decision. With this participant removed, the correlation between Shipley vocabulary and word frequency was marginally significant, r = .174, $p = .054$.

participants showing positive betas and some showing negative betas, but this effect seems to reflect small inhibitory valence effects for participants who read less, with a tendency towards small facilitatory valence effects for participants who read more.

Accuracy. As noted earlier, accuracy is reported but should be interpreted cautiously. Accuracy analyses were conducted by considering correct versus incorrect trials, excluding any RT or *z*-score outliers or microphone errors. B weights were derived from binary logistic regression analyses conducted on the individual subject level for each variable in each task. The correlation between age and accuracy word frequency B weights (see Table 9) were not significant for animacy judgment or lexical decision, but the word frequency effect in naming showed a significant positive relationship with age, $r = .223$, $p = .017$ (see Fig. 9), and the quadratic effect of age, $r = .241$, $p = .01$.

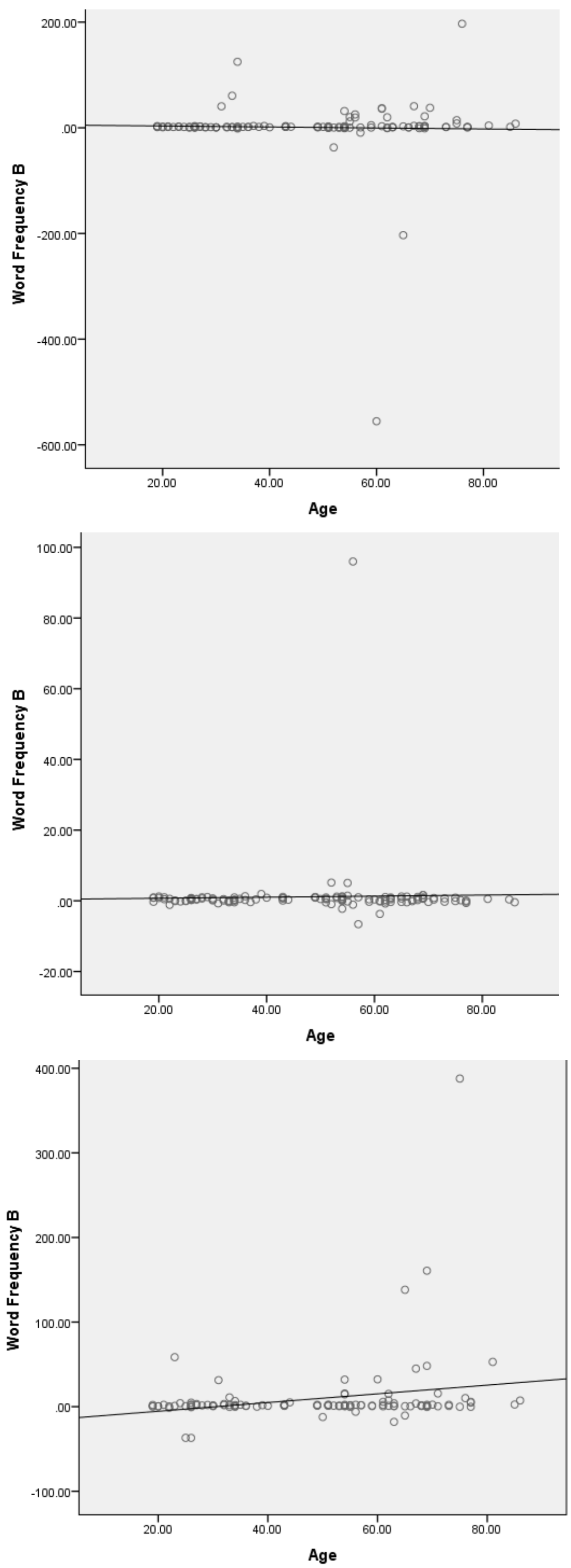

Figure 9. Linear and Quadratic Effects of Age on Word Frequency Accuracy Betas, Animacy Judgment (top), Lexical Decision (middle), and Naming (bottom)

The significant relationship between age and word frequency B in the naming task was not observed when any control variables were accounted for. However, significant or marginally significant positive correlations between age and word frequency B for animacy judgment did arise when Short Blessed Test score (and quadratic Short Blessed Test score), vision quadratic, education (and education quadratic), vocabulary (and vocabulary quadratic), and reading per week (and reading per week quadratic) were entered as control variables (see Table 10).

The last comparison of interest is correlations between other predictor variables and demographic variables (Table 11). Accuracy showed several significant correlations at the $p < .01$ level. First, age was associated only with length in the naming task, a positive association reflecting larger length effects as age increases. Short Blessed test was negatively associated with length in the animacy judgment task and concordance in the lexical decision task, and was positively associated with concreteness and consistency in the lexical decision task, but these relationships were potentially driven by an outlier with a high Short Blessed Test score.

3.5 Linear Mixed Effects Modeling

Linear mixed effects modeling was conducted in R using the LMER function (Baayen, Davidson, & Bates, 2008). This type of analysis allows simultaneous modeling of trial-, item- and subject-level effects, as well as exploration of the influence of a random intercept for subject and target item. Here all data were analyzed, with task as a factor entered along with the standard set of predictors (phonological onsets, length in letters, word frequency, concordance, consistency, valence, and concreteness) as fixed effects, and subject and target item with random intercepts. Furthermore, the critical Age × Word Frequency interaction was included, along with Age × Word

Frequency × Task, and all of its 2-way interactions[7]. Correct trials with *z*-scores inside of 3 SDs of each participant's mean were used (see filtering details above), and the dependent variables were raw RT and accuracy. Raw RT was used instead of *z*-scores because the random intercept for subject is another way to account for general slowing that occurs across the lifespan.

[7] The critical Age × Word Frequency interaction was also assessed by building a basic model (length, word frequency, concordance, consistency valence, concreteness, prior four trial RTs, and overall trial number), then a second model with the Age × Word Frequency × Task interaction added, plus all of its 2-way interactions. Although adding these interactions increased fit statistics over the basic model, χ^2 (6) = 6304, $p < .001$ for raw RT, and χ^2 (6) = 11985, $p < .001$ for log-transformed RT, the Age × Word Frequency and the Age × Word Frequency × Task interactions were non-significant, *p*s > .23.

Table 12.
Linear Mixed Effects Model Betas for RT and Accuracy

	RT	Log-Transformed RT	Accuracy
Length	.061***	.075***	.029
Word Freq	-.227***	-.184***	.197***
Concordance	-.047***	-.042***	.539***
Consistency	-.002	-.004	.046
Valence	.004	.002	.051*
Concreteness	-.032***	-.035***	-.001
Age	.212***	.230***	.484***
Task	-.585***	-.677***	.640***
Age × Task	-.058***	-.05***	-.071**
Word Frequency × Task	.045***	.016***	.254***
Age × Word Frequency	-.003	.003	-.074
Age × Word Frequency × Task	.004	.004	.020

Note. + p < .10. * p < .05. ** p < .01. *** p < .001.

Raw RTs. Table 12 displays the results from the linear mixed effects analysis. Phonological onsets were included in the analyses but are not displayed in the table above. In this analysis, the Age × Word Frequency was not significant, p = .657, nor was the Age × Word Frequency × Task interaction, p = .199. Therefore this analysis indicated that there was no relationship between age and word frequency effects overall, or in any task individually. An additional analysis added participant characteristics (Short Blessed test score, vision, education, OSpan, Shipley vocabulary, and hours read per week) one at a time to the model to assess the influence of accounting for these variables. In each case, the Age × Word Frequency and Age × Word Frequency × Task interactions remained non-significant, except that accounting for OSpan led to a significant Age × Word

Frequency interaction, $p = .049$, and a marginally significant Age × Word Frequency × Task interaction, $p= .051$. To explore these interactions, separate mixed effects models were run for each task. This resulted in a significant Age × Word Frequency interaction for lexical decision reflecting larger word frequency effects for older adults, and a significant Age × Word Frequency interaction for naming reflecting just the opposite (see Fig. 10). Interesting, this is the opposite of what was found in Balota et al. (2004); in their study, older adults showed a slightly larger word frequency effect for the naming task and the young adults showed a larger word frequency effect for the lexical decision task. Their overall analysis (linear regression) and metric (change in R^2 when word frequency was added to the model) were both different than the current linear mixed effects modeling, but this in an intriguing contrast of obtained effects.

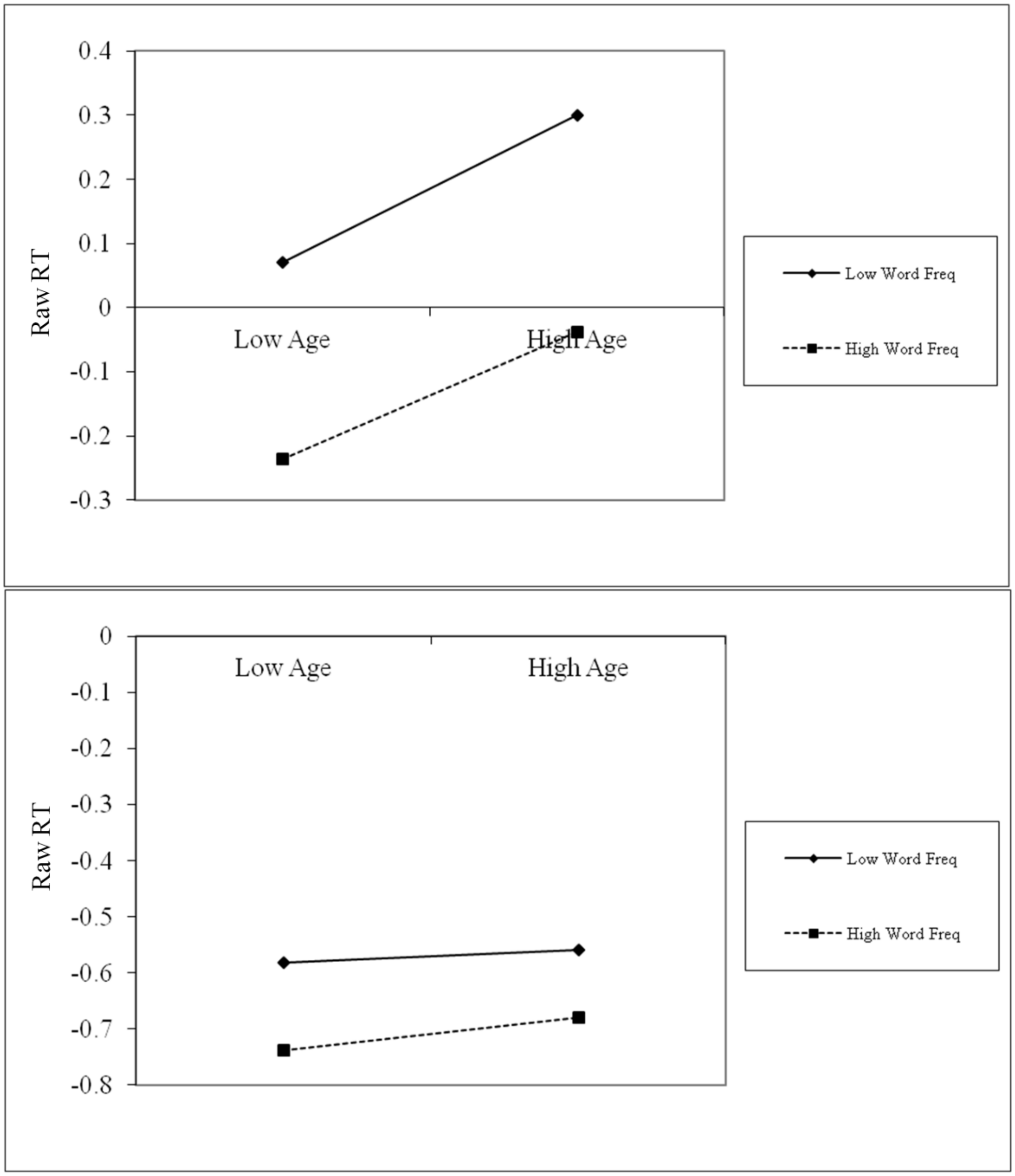

Figure 10. Linear Mixed Effects model interactions when controlling for OSpan, in lexical decision (top) and naming (bottom).

Finally, because of the emerging importance of vocabulary in the current dissertation, it was explored more fully in the linear mixed effects models. An Age × Word Frequency × Task × Vocabulary interaction term was added to the model, along with all lower-order interactions. The four-way interaction was not significant, $p = .802$, but the Word Frequency × Task × Vocabulary interaction was, $p < .001$. All tasks showed the predicted attenuation of word frequency effects as vocabulary increased, albeit to different magnitudes (e.g., an already-modest word frequency effect in the naming task was nearly eliminated at the highest vocabulary level; see Fig. 11).

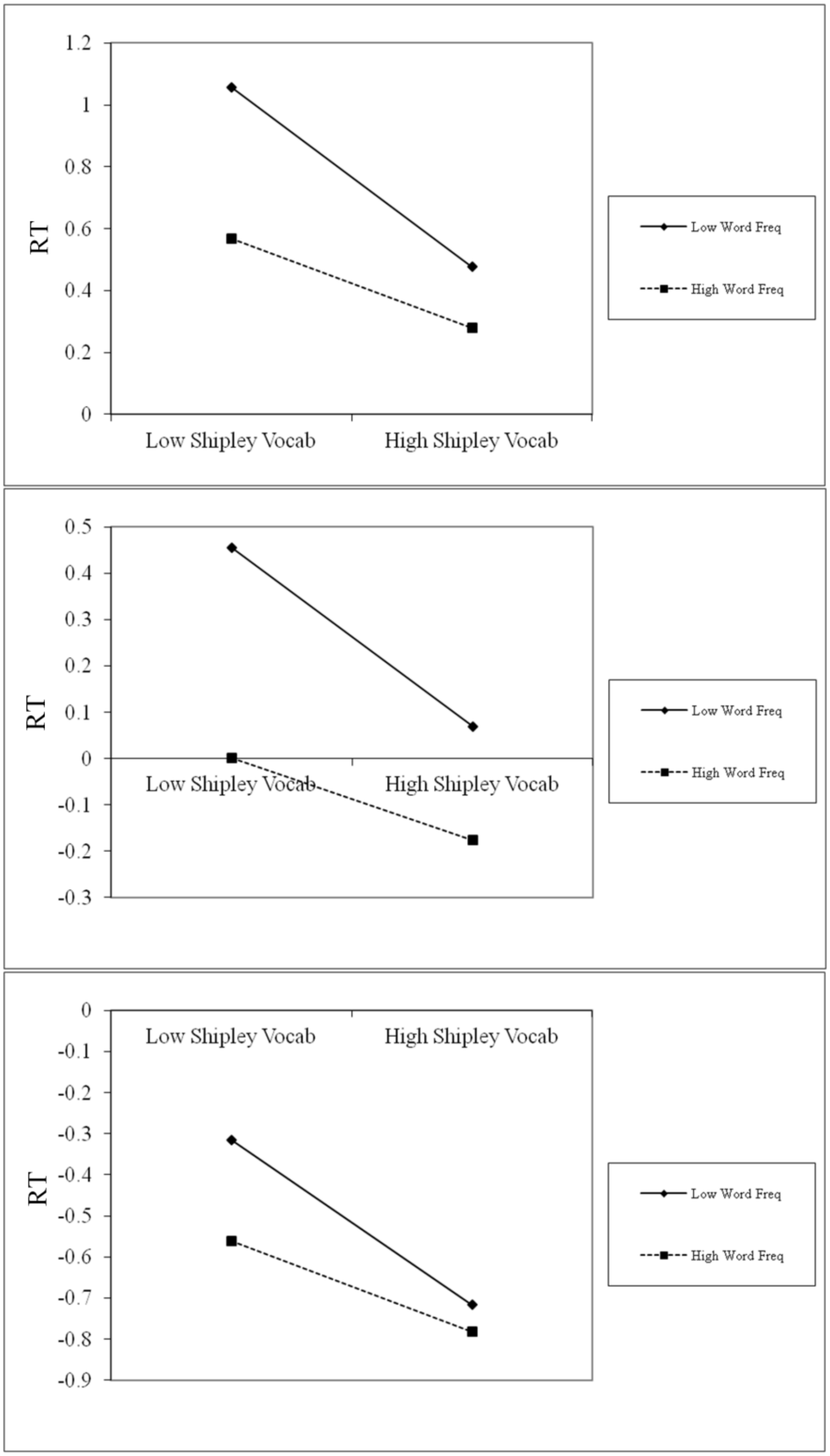

Figure 11. Word Frequency × Vocabulary Interaction for animacy judgment (top), lexical decision (middle), and naming (bottom).

Log-Transformed RTs. The same analyses were performed with log-transformed RT because raw RT is positively skewed (see also Table 12). These analyses mirrored the raw RT analyses; the Age × Word Frequency interaction was not significant, $p = .592$, nor was the Age × Word Frequency × Task interaction, $p = .128$. These transformations are common in mixed effects modeling but can reveal spurious interactions (see Balota, Aschenbrenner, & Yap, 2013). Fortunately, in the current study both produced the same results with respect to the critical interaction.

Accuracy. Accuracy was examined using the binomial family option of the LME function (see also Table 12). The Age × Word Frequency and Age × Word Frequency × Task interactions were not significant, $p > .10$, so no additional analyses were performed.

Additional Considerations. Conducting linear mixed effects analyses allows examination of additional factors which may influence responding. Full consideration of these factors is outside the scope of the dissertation, but it is important to explore at least whether these factors interact with the Age × Word Frequency or Age × Word Frequency × Task interactions.

First, the effect of trial number was assessed to see how performance unfolds across a task. One might predict overall fatigue or practice effects, the latter causing overall slowing and the former overall speeding, across trial number. This was borne out in the analyses, as trial number produced a significant main effect, $\beta = .012$, $p < .001$, reflecting slower RTs as the tasks progressed. However, including trial number as a factor did not change the influence of any other variables, nor did it reveal a significant Age × Word Frequency or Age × Word Frequency × Task interactions, *p*s < .191.

Second, the influence of the prior four trial RTs was assessed, each prior trial individually as well as all four prior RTs together. Four prior trials was chosen on the basis of computational limitations (i.e., with so many predictors, adding any more than four prior RTs caused program termination). Trial history has been of recent interest, from the perspective that participants may be sensitive to the difficulty and speed of prior trials and modulate processing accordingly (e.g., Adaptation to the Statistics of the Environment model, Kinoshita, Mozer, & Forster, 2011). Some studies have found prior trial effects on performance (Masson & Kliegl, 2013, although see O'Malley & Besner, 2013, for a failure to find an influence of prior trial on naming task performance, and Balota et al., 2013, for the perspective that these effects can be an artifact of RT transformations). In the current dissertation, the main effects of the prior four RTs each separately and all together showed significant and positive main effects, ps < .001, but the Age × Word Frequency and Age × Word Frequency × Task interactions still failed to reach significance, ps > .228 (and, like trial number, including trial number did not change the influence of any other variables). In every case, prior trial RTs were positively associated with current trial RTs, and this influence became weaker with more distance from the current trial (β = .169 for one trial back, β = .143 for two trials back, β = .134 for three trials back, and β = .128 for four trials back when entered separately, and β = .133 for one trial back, β = .091 for two trials back, β = .091 for three trials back, and β = .082 for four trials back when entered into one model all together, all ps < .001).

3.6 Task Specificity of Predictor Variables

As discussed above, tasks are often assumed to be process-pure (Jacoby, 1991). As a result, much of the prior work in visual word recognition lacks consideration of the task-specific processes brought online by task demands (see Balota & Yap, 2006). Having participants in the current study complete three tasks with the same stimuli allows for direct analysis of task-general and task-specific influences. One targeted question along these lines concerns age differences in the extent to which participants modulate performance as a function of task. Studies have examined this issue in young adults by looking at how the significance and relative strength of predictors changes across tasks (Andrews & Heathcote, 2001; Schilling, Rayner, & Chumbley, 1998; Yap, Pexman, et al., 2012). Because task-specific modulation of performance, such as biasing dimensions of the stimulus that are likely to be relevant for the task at hand, or inhibiting dimensions of the stimulus unlikely to be helpful for the task at hand, involve attentional control, one might predict that older adults show less task-specific modulation of performance. There is a hint of this in Balota et al. (2004), in which older adults' naming and lexical decision RTs are more highly correlated than young adults (R^2 = .08 for young, .17 for old). Alternatively, one might expect some preservation of this attentional control because the tasks are language based, and the stimulus dimensions are more highly familiar to older adults than young (Jenkins et al., 2000). If older adults' performance is not as influenced by task demands, this provides additional evidence for attentional breakdowns, even within the domain of language.

Table 13.
Inter-Task Correlations by Variable

	AJT & LDT	AJT & NMG	LDT & NMG
Overall *z*-scores	.233*	-.077	-.047
Length	.000	.024	.029
Word Freq	.207*	.345***	.152+
Concordance	.140	-.081	-.077
Consistency	.127	-.124	.140
Valence	-.060	.020	.114
Concreteness	-.043	.044	.014

Note. AJT = Animacy Judgment Task, LDT = Lexical Decision Task, NMG = Naming Task, + $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

In order to address this issue, correlations among the subject-level beta weights for the three tasks were assessed. Correlations between animacy judgment and lexical decision, animacy judgment and naming, and lexical decision and naming provide information on the extent of task-specific and task-general processing. These correlations were run for overall *z*-scores, as well as for *z*-scores for each variable separately (see Table 13). Surprisingly, only a few predictor variables were correlated across tasks. Specifically, the word frequency betas were significantly or marginally correlated with one another, animacy judgment and lexical decision, r = .207, $p = .022$, animacy judgment and naming, r = .345, $p < .001$, and lexical decision and naming, r = .152, $p = .095$. It is striking that so few correlations among task effects were obtained. It is possible that this reflects error variance in the estimates. This is unlikely, however, given the high reliability of these predictor variables within each of the specific tasks. It is more likely that the few correlations among tasks reflects the task specific processing engaged for each. Only word-frequency produced reliable effects, but these effects were not robust. Finally, there is no evidence for changes in the task-specific processing as a function of age, because partialling out age and computing correlations among tasks did not modulate the observed correlations (see Table 14).

Table 14.

Inter-Task Correlations by Variable, Partialled by Age

	AJT & LDT	AJT & NMG	LDT & NMG
Overall *z*-scores	.214*	.024	-.013
Length	.002	-.021	.030
Word Freq	.207*	.342***	.152+
Concordance	.139	-.072	-.076
Consistency	.124	-.130	.139
Valence	-.064	.020	.119
Concreteness	-.043	.044	.012

Note. AJT = Animacy Judgment Task, LDT = Lexical Decision Task, NMG = Naming Task, + $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

3.7 Analyses Split by Animacy

One might expect differing influences of variables, particularly word frequency, on animate and inanimate words. If participants are framing the animacy judgment in terms of animacy as the reference category, then animate words might receive a small boost in activation or priming, decreasing the word frequency effect (Balota & Chumbley, 1984). To assess the influence of animacy status in the current study, animacy (living versus non-living) was entered as a variable in the overall item-level regression analyses. The main effect of animacy was significant for animacy judgment only, $p < .001$, but the interactions of several variables with animacy were significant for all three tasks (see Fig. 12).

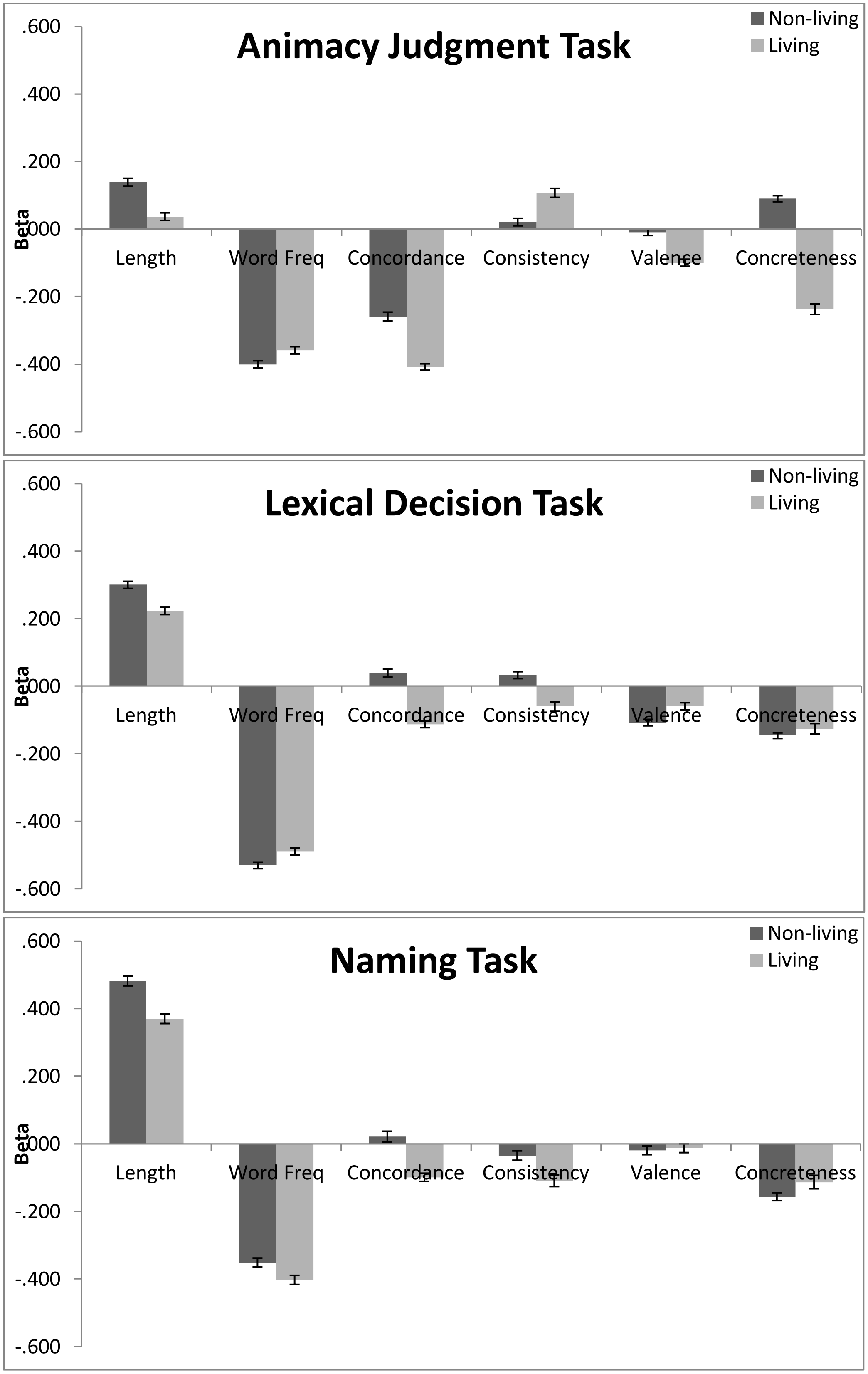

Figure 12. Item-level analyses split by animacy status (living versus non-living).

Perhaps not surprisingly, since living and non-living words differed on mean concordance, the influence of concordance differed significantly or marginally across all three tasks. For all three tasks, living words showed a significant facilitatory effect of concordance, and non-living words showed a smaller (animacy judgment) or null (lexical decision, naming) effect of concordance. This is possibly due to the fact that the words characterized as living had lower concordance scores; these words were harder, and slower, to characterize in the animacy judgment task. There were also other variables which produced interactions with animacy, including length and concreteness for animacy judgment (and valence, marginally). The Length × Animacy interaction reflected robust length effects for non-living words only, whereas the Concreteness × Animacy interaction reflected facilitatory effects for living words but inhibitory effects for non-living words. The marginal Valence × Animacy interaction reflected valence effects only for living words. Lexical decision showed only a marginally significant Word Frequency × Animacy interaction, which reflected smaller effects for living words. Finally, there was a significant Length × Animacy interaction for naming, which reflected smaller length effects for living words. These findings are not consistent with the prediction that participants are framing the animacy judgment task as a "living judgment", as this would result in smaller word frequency effects for living words in the animacy judgment task, whereas here the effect was in the lexical decision task[8].

[8] However, when these analyses were explored in a split by age group, the predicted Word Frequency × Animacy interaction (i.e., smaller word frequency effects for living words) did appear for younger and older adults, but a reversal occurred for middle-aged adults, explaining the non-significant overall interaction. Furthermore, the observation from other analyses that age does not modulate the word frequency effect seemed to hold for both living and non-living words.

Chapter 4: General Discussion

This dissertation examined word frequency, a hallmark of visual word recognition, and its influence on performance across the lifespan. In spite of the prevalence and robustness of word frequency effects in nearly any task relating to language, there is little consensus within the field as to the underlying mechanisms producing the word frequency effect, and there is relatively little consideration of the task differences and individual differences characteristics which modulate the influence of word frequency on performance. The current dissertation considered task differences and individual differences including, most critically, age.

All models of visual word recognition make assumptions regarding the role of word frequency, so the dissertation project afforded a unique opportunity for adjudication among them. In particular, models differ on predictions for the influence of age on the word frequency function. As described in the Introduction, the full spectrum of potential influences is predicted by various models, including a smaller, larger, or equal word frequency effect with increasing age. Past studies that have investigated the word frequency effect in young and older adults have produced varying patterns of effects. This difference in pattern of results is potentially due to task differences, stimulus selection issues, and potential participant differences, all of which were controlled or examined in the current study. The word frequency effect across three word recognition tasks was measured across a large set of participants and across a large set of stimuli. In addition, instead of comparing college freshman and sophomores to select samples of older adults, participants were drawn from a diverse pool and potentially relevant participant characteristics were directly measured.

The primary finding was a lack of association between word frequency and age. That is, the word frequency effect does not appear to change across the lifespan, at least in response times. This was supported first by the subject-level analyses, which were best poised to address this question. In those analyses, age and word frequency were considered continuously and correlations were computed between the participants' word frequency coefficients and their age. There was no significant correlation between the two. This finding was mirrored in the item analyses with age trichotomized, which showed no difference between age groups' word frequency betas in the animacy judgment task. The third convergence on this lack of age-related change in word frequency effects came from the linear mixed effects modeling, in which trial-, item-, and subject-level data were considered simultaneously. No significant age by word frequency or age by word frequency by task interactions were obtained. The lack of significant interactions persisted whether raw response times or log-transformed response times were considered, and when potential confounds were added to the model (e.g., prior four trials RTs, overall trial number).

As emphasized in the Introduction, one important potential explanation for the inconsistency in prior literature is participant individual difference characteristics which modulate the word frequency effect and which differed across prior studies. For example, vocabulary has been shown to differ across younger and older adults (Verhaeghen, 2003), and to influence the word frequency effect (Chateau & Jared, 2000). Therefore it may be individual differences characteristics associated with age and word frequency which are driving the modulations of the age by word frequency interaction across different studies, and the lack thereof in the current study. The current study was able to investigate the influence of participant characteristics on the word frequency/age relationship. Measures of general cognition (Short Blessed Test score), vision (factor score computed from near and far, high- and low-contrast tests), education (number of years of school), working memory ability (operation span), vocabulary (Shipley vocabulary number correct), and reading experience (self-reported number of hours read per week) were all collected. The influence of these characteristics was assessed in two ways. First, partial correlations between age and word frequency betas on the subject level were computed controlling for these participant characteristics, one at a time and with two targeted combinations (vocabulary, vision, OSpan, or vocabulary, vision, education). None of the partial correlations between age and word frequency betas reached significance except that controlling for vision and the quadratic effect of vision both led to a significant correlation between age and the word frequency effect in the naming task. This reflected a *smaller* word frequency effect with increasing age and, intriguingly, this occurred despite specifying normal or corrected-to-normal vision as an inclusion criterion for the study. If this correlation is not spurious, vision seems to contribute to modulation of word frequency effects; participant differences with respect to vision seem to have downstream influences on processing. There was also a marginally significant correlation between age and the word frequency effect in the animacy judgment task when controlling for vocabulary and the quadratic effect of vocabulary. This marginal partial correlation actually reflected the opposite, a larger word frequency effect with increasing age. This analysis is especially interesting because of the robust correlation between age and Shipley vocabulary—in studies such as Balota and Ferraro (1996), Ratcliff et al. (2004), and Spieler and Balota (2000) vocabulary was equal across younger and older adults and they found an increasing word frequency effect with increasing age[9].

The second way participant individual differences characteristics were explored in the current study was by adding them one at a time to a linear mixed effects model. This procedure did not change the lack of observed age by word frequency or age by word frequency by task interactions. The only exception to this was adding OSpan to the model the age by word frequency

[9] However, see Bowles & Poon (1981) and Whiting, Wythe, Madden, Langley, Denny, Turkington, et al. (2003), whose younger and older participants were equal on vocabulary and showed ostensibly equal word frequency effects (the former in double lexical decision task, the latter in a standard lexical decision task). However, both studies did show trends of higher word frequency effects for older adults.

by task interaction became significant. This reflected small but significant age by word frequency interactions in lexical decision, in which word frequency effects increased with age, and naming, in which word frequency effects decreased with age.

Taken together, the results from the present study yielded little evidence of a change in the word frequency effect as a function of age. There are some violations of this pattern but they are small, at times marginally significant, and difficult to interpret with respect to theory. One possibility is that there are small, subtle changes in the word frequency effect as a function of age caused by participant individual differences characteristics. A second possibility is that there are some spurious findings because of the number of comparisons in the current dissertation. In fact, if a *p*-value correction were imposed, none of the partial correlations between age and word frequency effects, or the linear mixed effects interactions with OSpan in the model, would be significant at even the $p < .01$ level.

Thus far discussion has focused on response time analyses. Patterns in accuracy turned out to be considerably more complicated. As is typical, accuracy was highly skewed in the current study in large part because participants were at ceiling performance, 95% or better. Despite this ceiling performance, significant differences were observed in the accuracy measures. Specifically, accuracy measures showed significant decline in word frequency effects as a function of age in the animacy judgment and lexical decision tasks, a relationship which holds up in nearly all of the partial correlations controlling for demographic characteristics. Furthermore, the item-level analyses with age trichotomized mirrored these patterns: word frequency effects decline as a function of age group in the animacy judgment and lexical decision tasks, and are more similar, or perhaps even increasing across age group, in the naming task. Accuracy as analyzed in the linear mixed effects modeling did not show significant age by word frequency or age by word frequency by task interactions. It is difficult to know how to interpret significant age by word frequency interactions in accuracy; prior studies have examined primarily RT, and models of visual word recognition have been primarily based on response latency data. Indeed, it is possible that age-related differences in speed-accuracy tradeoffs may play an important role (see Starns & Ratcliff, 2010). For future work in this area, one may need a model that simultaneously captures both reaction time and accuracy (e.g., the Diffusion Model), but this may need to involve different paradigms or manipulations that afford accuracy estimates off of ceiling for individuals.

4.1 Reconciling Inconsistencies in Prior Literature

One goal of the current dissertation was to examine several characteristics of prior studies to see if they influence the age by word frequency interaction. One potential source of inconsistency was participant characteristics, described above. The current study attempted to reconcile this by measuring participant characteristics and recruiting a fresh set of diverse participants, compared to other studies which used college students and healthy community-dwelling retirees (both populations who participate very often in psychological studies). This was critical because it is possible to obtain different signatures of processing when distinct populations are used (even in cases of two different college student populations, as in Yap, Tse, & Balota, 2009) Participant characteristics as assessed here do not appear to strongly modulate the pattern of obtained results. One small but intriguing finding was that the effect of partialling out vocabulary in the current study somewhat mirrors other studies with younger and older adults matched on vocabulary. Specifically, several studies which find increasing word frequency effects with age included younger and older adults who are matched on vocabulary (Balota and Ferraro, 1996; Ratcliff et al., 2004; and Spieler and Balota, 2000; in comparison to the typical age superiority in vocabulary scores, Verhaeghen, 2003). In the current study, controlling for vocabulary led to a marginally significant association between age and word frequency in the animacy judgment task. Although this connection is tenuous (marginal significance, and in the task less examined in prior literature), it is intriguing.

Another potential source of inconsistencies in prior literature is the task used to draw inferences. Clearly task had a large influence on the main effects of variables; word frequency differed by 0.16 for item-level *Z*-score betas, 0.19 for accuracy betas. Most theoretical accounts lack consideration of task-specific influences and few studies employ more than one task. Across most analyses in the current dissertation there did not seem to be an age by word frequency interaction in <u>any</u> task. However, tasks did show some subtle differences with respect to the age by word frequency interaction, e.g. when participant characteristics were controlled for and the interaction was significant in naming but not the other tasks. Studies such as the current one highlight the importance of using multiple tasks to triangulate task-general and task-specific processes, and from which to build theories and models (Grainger & Jacobs, 1996). One important note is that some of the literature reviewed in the Introduction included eye-tracking, and even full text processing tasks. The current study may diverge from those results not because of methodological differences between studies but because the domain of those hypotheses is slightly different (full text processing versus single word recognition). However, the hypotheses advanced in those studies are arguably amenable to study in the current dissertation, e.g., that older adults appeal to a “partial-reading” strategy as a result of visual decline and increased experience in visual word recognition as well as full text processing.

The last potential source of inconsistencies between the present study and the prior studies is that the methods of considering age differ. Past studies have primarily used an extreme groups approach, i.e., dichotomizing young (college students) versus older individuals (often past 60).

However, to the extent that extreme groups could be explored in the present analyses, there seemed to be little influence of this difference. Specifically, subject analyses for the most part mirrored the trichotomized item analyses. Another issue relevant to age considerations is the age spectrum recruited in the current study. Although this tactic had many benefits, as described in the Introduction, it may also have resulted in less power because of smaller numbers of participants recruited than in the more typical extreme-groups design (e.g., 18-25 years old, 60-75 years old). In fact, the current study included only 14 participants ages 18-25, so some of the disparate results may be that the younger adult group was not as well-represented or as young as prior studies (although importantly, 29 participants ages 18-30 were included in all trichotomized age analyses).

Importantly, the current study also provided some interesting insights into middle-aged adults' visual word recognition processing, which has been relatively unexamined. This is important because of the assumption of prior work that the effect of interest has a linear relationship with age, e.g., middle-aged adults would show intermediate effects between younger and older adults' (Salthouse, 2000). This is not always the case for cognitive processing, although sampling continuously is not often undertaken (Verhaeghen & Salthouse, 1997). In fact, middle-aged adults in the current study looked qualitatively different than younger and older adults in some cases. One intriguing pattern along these lines was the (marginally significant) quadratic effect of age in naming word frequency betas, such that middle-aged adults showed larger word frequency effects than younger and older adults. This pattern also appeared in lexical decision and naming word frequency effects for the item analyses trichotomized by age; word frequency effects were largest for middle-aged adults instead of their means falling linearly between younger and older adults.

4.2 What Model Can Account for the Observed Effects?

A primary thrust of the dissertation was to assess the observed results with respect to model predictions. First, a major concern is that models and theories of visual word recognition do not account for task differences. This may be because studies often do not include more than one task from which to theorize or build models. The current study provides a step forward in theory and modeling by comparing tasks. In the one prior study which included more than one task (Balota et al., 2004), overall task performance (overall *z*-scores in lexical decision and naming) was more highly correlated for older adults than younger adults. This was taken as evidence that task-specific processing is more difficult for older adults. This hypothesis not supported in the current study; correlations among the overall *z*-scores and, even more specifically, the item predictor beta weights for different tasks did not vary as a function of age. Another hypothesis with respect to task is that one might expect different age by word frequency interactions for each task; e.g., older adults may rely disproportionately on frequency (accrued more throughout lifetime), so there should be larger age changes in a the more frequency-dependent lexical decision task than the other two tasks. This idea is not supported in the simple correlations between age and word frequency betas, nor in the basic linear mixed effects modeling (no significant age by word frequency by task interaction). In fact, tasks do start to pull apart when partialling out subject variables of interest in subject-level and LME analyses, but not in this manner. Specifically, when vision or OSpan is controlled for (in the subject and LME analyses, respectively), a correlation between word frequency betas and age emerged (reflecting smaller word frequency effects with increasing age).

Across multiple analysis techniques, the primary finding in the current study was *similar* word frequency effects for participants across the lifespan. This prediction is consistent with the rank frequency account (Murray & Forster, 2004), which posits that rank frequency of an item, not absolute frequency, is predictive of performance. The rank frequency of items should not change across the lifespan, only the absolute frequency, so one should see equivalent word frequency effects across the lifespan.

Constant word frequency effects across the lifespan are inconsistent with a logogen-type account (e.g., Morton, 1969) because an unembellished version of this account posits increasing exposure to words reaching eventual asymptote. The notion is that young adults already have many high frequency words at asymptote. However, as one ages more low-frequency words would also reach asymptote. Hence, this model would predict a decreasing word frequency effect in older adults. There was relatively little evidence in support of this hypothesis. Because a logogen-type account does not specifically make predictions about age, one might posit additional age factors which work in opposition to and age-related change in the word frequency effect. For example, increasing experience on mostly high-frequency words drives performance towards a larger word frequency effect with age, but greater vocabulary knowledge, or greater benefit of experience for low-frequency words since they are further from the performance asymptote, both drive performance towards a smaller word frequency effect with age.

The lack of age by word frequency effect is also inconsistent with a Transmission Deficit Hypothesis, as it posits that weakened transmission of activation throughout the cognitive system is disproportionately large for infrequently-accessed words relative to more frequently-accessed words. It is also inconsistent with theory that lexical representations become more holistic over time, e.g., Samuels, LaBerge, & Bremer (1978) and Spieler & Balota (2000). However, there is some evidence for the other side of this theory positing disruptions in sublexical or piecemeal word activation with increasing age in the form of the reduced length effects observed in certain aspects of the current study (e.g., Allen et al., 1991, 1995, 2011).

4.3 Limitations of the Current Study

There were some limitations to the current study which limit statistical or interpretive power. Because of the broad goals of the current study, including recruitment across the lifespan and three tasks instead of the more typical one, there is less power than some prior studies (e.g., Balota et al., 2004). Furthermore, as in any study of visual word recognition, the current study included what are potentially imperfect operationalizations of item predictor variables. Specifically, only first phonemes were considered, and onsets were coded categorically, based on presence or absence of a feature rather than the specific feature present. This means, for example, that all voiced versus unvoiced onsets were treated the same which is a potentially faulty assumption. Future studies may include a more nuanced consideration of phonological onsets (as in Kessler et al., 2002). A similar problem exists for the consistency variable and may in part explain the small effects of that measure. That is, defining consistency in terms of only the syllable rime ignores systematic variation caused by context, including the onset of that syllable and other syllables in the word. Consistency is a more complex construct for multisyllabic words than it is for monosyllabic words, although reliable effects of consistency have been found in a disyllabic or multisyllabic dataset using a similar consistency calculation or considering consistency categorically (e.g., Jared, et al., 1990; Chateau & Jared, 2003; Yap and Balota, 2009). Another set of limitations are related to the analytic approaches. Item analyses involved trichotomizing a continuous variable, a practice which is generally advised against (Cohen, 1983). Subject and mixed effects analyses seemed to support the null hypothesis, which is always difficult to "prove". This issue was mitigated somewhat by employing three converging analytic techniques with multiple dependent variables, and for the most part these analyses converged. Last, linear mixed effects modeling may inflate type I error rate (Barr, Levy, Scheepers, & Tily, 2013), and RT transformations may produce spurious effects (Balota et al., 2013). However, the primary finding was of no interaction between age and word frequency, and so cannot be either a type I error or a spurious effect.

4.4 Future Directions

Despite the limitations in the current study, there were several strengths of the current study. It recruited a large, diverse subject pool of mostly first-time participants who were more representative of the population, including middle-aged adults. The current study also employed multiple tasks, converging analyses, and a large stimulus set. This resulted in a rich dataset with potential for future exploration, as well as targeted follow-up studies. Future studies may include a fuller consideration of other item predictor variables and their relation to age (e.g., length and orthographic N are theoretically important, as in Spieler & Balota, 2000), as well as a fuller characterization of the relatively under-studied middle age group. Future analyses may explore variables associated with word frequency, such as age of acquisition (e.g., Morrison et al., 2002), contextual diversity (Adelman et al., 2006), cumulative word frequency (Caza & Moscovitch, 2005; Morrison et al., 2002), and frequency trajectory (Zevin & Seidenberg, 2002). Age comparisons would be especially interesting here since many of these variables specifically implicate age predictions. Of course, as the number of predictor variables increases so do concerns about collinearity of variables, so taking another large-scale approach is likely necessary.

An interesting extension of the current work is using it to examine computational models directly, as in Spieler & Balota (1997). This technique of "bringing models down to the item level" (Spieler & Balota, 1997; Yap & Balota, 2009) allows more detailed model assessment by comparing participant response times or accuracy to model settling times or error scores on an individual item basis. This allows direct assessment of nuances in the data on top of broad categorical patterns. One may examine several computational models in this manner, including the Connectionist Dual Process Model (CDP++, Perry, Ziegler, & Zorzi, 2010), the Junction Model (Kello, 2006), and the Bayesian Reader (Norris, 2006), which have implementations including English multisyllabic words. Examining these models with the current data may give rise to further predictions regarding age, word frequency, or other variables. Similarly, the models may benefit from a consideration of lifespan data, as they vary on their consideration of early learning mechanisms.

Another future direction includes using advanced statistical techniques regarding other aspects of the data. For example there is currently some debate as to the form of the function relating RT to word frequency (see Adelman & Brown, 2008; Murray & Forster, 2004; Norris, 2006). Models of visual word recognition predict different functions (e.g., logarithmic, power, or exponential). Therefore another exploration of this factor and whether it is modulated by age would be beneficial. RT distributional analyses, which look beyond mean RT to include measures of the standard deviation and skewed tail of the distribution, may be useful. There is evidence that task, word frequency, and age, as well as other participant characteristics, influence different components of the RT distribution and may provide leverage on questions of changes in processing with age. Another advanced technique that would go beyond mean RT is using principal components analysis to reduce item and subject variables (like the vision factor here, but applied more broadly, e.g., by Yap et al., 2012, who found a lexical/word frequency factor, a structural factor, and a neighborhood factor in item variables). Using principal components may provide more stable estimates of item and subject factors while appropriately handling large and potentially problematic multicolinearity of variables.

One important conceptual aspect of the current study is the challenge to the "static" model of visual word recognition, which many theories and models implicitly assume. That is, existing models and theories are based on an adult visual word recognition system with no mechanism for development or change (with some exceptions, e.g., Plaut et al., 1996; Seidenberg & McClelland, 1989). The current study suggests some nuanced ways the system may change even from middle to older age, so it seems a natural extension to consider these issues in children. This would be an especially interesting group to look at, because although their visual word recognition is slower and less skilled than adults', it may involve qualitatively different processing (e.g., more component and less holistic processing; Samuels, LaBerge, & Bremer, 1978).

4.5 Conclusions

The present dissertation extended previous work on aging visual word recognition by using several tasks in order to study the impact of task demands, including a diverse lifespan participant sample, and examining the influence of participant characteristics. The consideration of participant characteristics such as vocabulary and working memory is critical in understanding age-related changes in visual word recognition (in particular the word frequency effect) and there is limited prior work on this issue.

The results provided evidence for stable word frequency effects across the lifespan. The finding of no interaction between age and word frequency persisted across multiple tasks (animacy judgment, lexical decision, and naming), analyses (subject-level, item-level, and linear mixed effects modeling), and dependent variables (*z*-scores, raw RT), as well as through the addition of participant characteristics (general cognition, vocabulary, education, hours read per week).

The exceptions to this observed pattern of age constancy in the word frequency effect were small deviations which were either marginal or in only a few analyses, except for a points of interest. First, controlling for vision led to smaller word frequency effects with age in the naming task. Second, controlling for vocabulary led to a (marginally significant) pattern of larger word frequency effects with age. Third, accuracy in the animacy judgment and lexical decision tasks showed a declining word frequency effect with age. The first two points are of theoretical interest, and the third is potentially in conflict with prior research but inspires further study.

Although there are some small changes in the word frequency effect as a function of age, the vast majority of analyses suggest age-independence. This pattern is best accommodated by Murray and Forster's rank order model (2004) in which exposure to words does not change their rank frequency, i.e., low-frequency words do not become more similar to high frequency words as a function of increasing experience. Interestingly, not only was there age constancy in the word frequency effect, there was also age constancy in the other variables studied (including length, valence, and concreteness). The present study also suggests that the changes that one finds beginning in the third decade of life in memory, attention, processing speed (Salthouse, 2004) do not extend to a language processing task in a variable-specific manner. In this light, the present results may support the contention that the language system has privileged protection against the onslaught of aging.

References

Abrams, L., & Farrell, M. T. (2011). Language processing in normal aging. In J. Guendouzi, F. Loncke, & M. J. Williams (Eds.), *The handbook of psycholinguistic and cognitive processes: Perspectives in communication disorders* (pp. 49-73). New York, NY: Psychology Press.

Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times.*Psychological Science*, *17*(9), 814-823.

Algom, D., Chajut, E., & Lev, S. (2004). A rational look at the emotional stroop phenomenon: a generic slowdown, not a stroop effect. *Journal of Experimental Psychology. General*, *133*(3), 323–338. doi:10.1037/0096-3445.133.3.323

Allen, P. A., Bucur, B., Grabbe, J., Work, T., & Madden, D. J. (2011). Influence of encoding difficulty, word frequency, and phonological regularity on age differences in word naming. *Experimental Aging Research*, *37*(3), 261–92. doi:10.1080/0361073X.2011.568805

Allen, P. A., Madden, D. J., & Crozier, L. C. (1991). Adult age differences in letter-level and word-level processing. *Psychology and Aging*, *6*(2), 261–71.

Allen, P. A., Madden, D. J., Weber, T. A., & Groth, K. E. (1993). Influence of age and processing stage on visual word recognition. *Psychology and Aging*, *8*(2), 274–82.

Allen, P. A., Wallace, B., & Weber, T. A. (1995). Influence of case type, word frequency, and exposure duration on visual word recognition. *Journal of Experimental Psychology. Human Perception and Performance*, *21*(4), 914–34.

Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? *Memory & Cognition*, *10*(6), 565–575. doi:10.3758/BF03202439

Andrews, S., & Heathcote, A. (2001). Distinguishing common and task-specific processes in word identification: A matter of some moment? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(2), 514–544. doi:10.1037//0278-7393.27.2.514

Augustine, A. A., Mehl, M. R., & Larsen, R. J. (2011). A positivity bias in written and spoken English and its moderation by personality and gender.*Social Psychological and Personality Science*, doi: 10.1177/1948550611399154.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390-412.

Balota, D. A., Aschenbrenner, A. J., & Yap, M. J. (2013). Additive effects of word frequency and stimulus quality: the influence of trial history and data transformations. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *39*(5), 1563–71. doi:10.1037/a0032186

Balota, D. A., & Chumbley, J. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(3), 340–357.

Balota, D. A., & Chumbley, J. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language*, *24*, 89–106.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283–316. doi:10.1037/0096-3445.133.2.283

Balota, D. A., & Duchek, J. M. (1988). Age-related differences in lexical access, spreading activation, and simple pronunciation. *Psychology and Aging*, *3*(1), 84–93.

Balota, D. A., & Duchek, J. M. (1991). Semantic priming effects, lexical repetition effects, and contextual disambiguation effects in healthy aged individuals and individuals with senile dementia of the Alzheimer type. *Brain and Language*, *40*(2), 181–201.

Balota, D. A., & Ferraro, F. R. (1993). A dissocation of frequency and regularity effects in pronunciation performance across young adults, older adults, and individuals with senile dementia of the Alzheimer type. *Journal of Memory and Language*, *32*, 573–592.

Balota, D. A., & Ferraro, F. R. (1996). Lexical, sublexical, and implicit memory processes in healthy young and healthy older adults and in individuals with dementia of the Alzheimer type. *Neuropsychology*, *10*(1), 82–95. doi:10.1037//0894-4105.10.1.82

Balota, D. A., & Yap, M. J. (2006). Attentional control and the flexible lexical processor: Explorations of the magic moment of word recognition. In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (pp. 229–258). Psychology Press.

Balota, D. A., Yap, M. J., & Cortese, M. J. (2006). Visual word recognition: The journey from features to meaning (a travel update). *Handbook of Psycholinguistics*, *2*, 285-375.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., … Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–59.

Balota, D. A., Yap, M. J., Hutchison, K. A., & Cortese, M. J. (2012). Megastudies: what do millions (or so) of trials tell us about lexical processing? In J. S. Adelman (Ed.), *Visual Word Recognition* (pp. 90–115). Hove: Psychology Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278.

Bowles, N. L., & Poon, L. W. (1985). Aging and retrieval of words in semantic memory. *Journal of gerontology*, *40*(1), 71-77.

Boyle, P. A., Yu, L., Wilson, R. S., Gamble, K., Buchman, A. S., & Bennett, D. A. (2012). Poor decision making is a consequence of cognitive decline among older persons without Alzheimer's disease or mild cognitive impairment. *PloS One*, *7*(8), e43647. doi:10.1371/journal.pone.0043647

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–90. doi:10.3758/BRM.41.4.977

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–11. doi:10.3758/s13428-013-0403-5

Burke, D. M., & MacKay, D. (1997). Memory, language, and ageing. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *352*(1363), 1845–56. doi:10.1098/rstb.1997.0170

Burke, D. M., & Shafto, M. (2008). Language and aging. In F. Craik & T. A. Salthouse (Eds.), *The Handbook of Aging and Cognition* (3rd ed., pp. 373–443). New York, NY: Psychology Press.

Carpenter, P. A., Miyake, A., & Just, M. (1994). Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 1075–1122). San Diego, CA: Academic Press.

Caza, N., & Moscovitch, M. (2005). Effects of cumulative frequency, but not of frequency trajectory, in lexical decision times of older adults and patients with Alzheimer's disease. *Journal of Memory and Language*, *53*(3), 456–471. doi:10.1016/j.jml.2005.03.005

Chae, B.C., Burke, D.M., & Ketron, P.P. (2002). Aging and picture naming: The role of phonological retrieval. Poster presented at the Cognitive Aging Conference, Atlanta, Georgia

Chateau, D., & Jared, D. (2000). Exposure to print and word recognition processes. *Memory & Cognition*, *28*(1), 143–53.

Chateau, D., & Jared, D. (2003). Spelling–sound consistency effects in disyllabic word naming. *Journal of Memory and Language*, *48*(2), 255-280.

Cohen, J. (1983). The Cost of Dichotomization. *Applied Psychological Measurement*, *7*(3), 249–253. doi:10.1177/014662168300700301

Colombo, L., Pasini, M., & Balota, D. A. (2006). Dissociating the influence of familiarity and meaningfulness from word frequency in naming and lexical decision performance. *Memory & Cognition*, *34*(6), 1312–1324. doi:10.3758/BF03193274

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. *Attention and Performance VI*.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–56.

Craik, F. (1994). Memory Changes in Normal Aging. *Current Directions in Psychological Science*, *3*(5), 155–158. doi:10.1111/1467-8721.ep10770653

Craik, F., & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F. Craik & S. Trehuti (Eds.), *Aging and Cognitive Processes* (pp. 191–211). New York: Plenum.

Cronin-Golomb, A., Gilmore, G. C., Neargarder, S., Morrison, S. R., & Laudate, T. M. (2007). Enhanced stimulus strength improves visual cognition in aging and Alzheimer's disease. *Cortex*, *43*(7), 952–966. doi:10.1016/S0010-9452(08)70693-2

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331.

Estes, Z., & Adelman, J. S. (2008a). Automatic vigilance for negative words in lexical decision and naming: comment on Larsen, Mercer, and Balota (2006). *Emotion (Washington, D.C.)*, *8*(4), 441–444; discussion 445–457. doi:10.1037/1528-3542.8.4.441

Estes, Z., & Adelman, J. S. (2008b). Automatic vigilance for negative words is categorical and general. *Emotion*, *8*(4), 453–457. doi:10.1037/a0012887

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin*, *125*(6), 777.

Forster, K. I. (1985). Lexical acquisition and the modular lexicon. *Language and Cognitive Processes*, *1*(2), 87-108.

Forster, K. I. (2000). The potential for experimenter bias effects in word recognition experiments. *Memory & Cognition*, *28*(7), 1109–15.

Forster, K. I., & Shen, D. (1996). No enemies in the neighborhood: Absence of inhibitory neighborhood effects in lexical decision and semantic categorization.*Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(3), 696.

Fozard, J. L., & Gordon-Salant, S. (2001). Changes in Vision and Hearing with Aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the Psychology of Aging* (5th ed., pp. 241–266). Academic Press, Inc.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always a means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787–814. doi:10.1016/j.jml.2007.07.001

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: the frequency-lag hypothesis. *Journal of Experimental Psychology. General*, *140*(2), 186–209. doi:10.1037/a0022256

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: a multiple read-out model. *Psychological review*, *103*(3), 518.

Hargreaves, I. S., & Pexman, P. M. (2012). Does richness lose its luster? Effects of extensive practice on semantic richness in visual word recognition. *Frontiers in Human Neuroscience*, *6*, 234. doi:10.3389/fnhum.2012.00234

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (pp. 193–225). New York.

Hino, Y., Lupker, S. J., & Pexman, P. M. (2002). Ambiguity and synonymy effects in lexical decision, naming, and semantic categorization tasks: Interactions between orthography, phonology, and semantics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 686–713. doi:10.1037//0278-7393.28.4.686

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 513–541.

James,C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology. Human Perception and Performance*, 1, 130–136. doi:10.1037/0096-1523.1.2.130

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language*, *29*, 687–715. doi:10.1016/0749-596X(90)90044-Z

Jenkins, L., Myerson, J., Joerding, J. A., & Hale, S. (2000). Converging evidence that visuospatial cognition is more age-sensitive than verbal cognition. *Psychology and Aging*, *15*(1), 157–75.

Katzman, R., Brown, T., Fuld, P., Peck, A., Schechter, R., & Schimmel, H. (1983). Validation of a short Orientation-Memory-Concentration Test of cognitive impairment. *The American Journal of Psychiatry*, *140*(June), 734–739.

Kello, C. T. (2006). Considering the junction model of lexical processing. In S. Andrews (Ed.), From inkmarks to ideas: Current issues in lexical processing. New York: Taylor and Francis.

Kessler, B., Treiman, R., & Mullennix, J. (2002). Phonetic Biases in Voice Key Response Time Measurements. *Journal of Memory and Language*, *47*(1), 145–171. doi:10.1006/jmla.2001.2835

Kessler, B., Treiman, R., & Mullennix, J. (2008). Feedback-consistency effects in single-word reading. In E. L. Grigorenko & A. J. Naples (Eds.), *Single-word reading: Behavioral and biological perspectives.* (pp. 159–174). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: a multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633. doi:10.3758/BRM.42.3.627

Kinoshita, S., Mozer, M. C., & Forster, K. I. (2011). Dynamic adaptation to history of trial difficulty explains the effect of congruency proportion on masked priming. *Journal of Experimental Psychology. General*, *140*(4), 622–36. doi:10.1037/a0024230

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*(1-2), 262–284. doi:10.1080/09541440340000213

Kramer, A. F., & Kray, J. (2006). Aging and Attention. In F. Craik & E. Bialystok (Eds.), *Lifespan Cognition: Mechanisms of Change* (pp. 57–69). Oxford University Press, USA.

Kučera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293-323.

Larsen, R. J., Mercer, K. A., Balota, D. A., & Strube, M. J. (2008). Not all negative words slow down lexical decision and naming speed: Importance of word arousal. *Emotion*, *8*(4), 445–452. doi:10.1037/1528-3542.8.4.445

Laubrock, J., Kliegl, R., & Engbert, R. (2006). SWIFT explorations of age differences in eye movements during reading. *Neuroscience and Biobehavioral Reviews*, *30*(6), 872–84. doi:10.1016/j.neubiorev.2006.06.013

Lovelace, E., & Twohig, P. T. (1990). Healthy older adults' perceptions of their memory functioning and use of mnemonics. *Bulletin of the Psychonomic Society*, *28*(2), 115–118.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203–208. doi:10.3758/BF03204766

Madden, D. J. (2007). Aging and Visual Attention. *Current Directions in Psychological Science*, *16*(2), 70–74. doi:10.1111/j.1467-8721.2007.00478.x

Masson, M. E. J., & Kliegl, R. (2013). Modulation of additive and interactive effects in lexical decision by trial history. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *39*(3), 898–914. doi:10.1037/a0029180

Mather, M., & Carstensen, L. L. (2005). Aging and motivated cognition: The positivity effect in attention and memory. *Trends in cognitive sciences*, *9*(10), 496-502.

McCoy, S. L., Tun, P., Cox, L. C., Colangelo, M., Stewart, R. A., & Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *The Quarterly Journal of Experimental Psychology*, *58*(1), 22–33. doi:10.1080/02724980443000151

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*(2), 165–178.

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they?. *Journal of Experimental Psychology: General*, *118*(1), 43.

Morrison, C. M., Hirsh, K. W., Chappell, T., & Ellis, A. W. (2002). Age and age of acquisition: An evaluation of the cumulative frequency hypothesis. *European Journal of Cognitive Psychology*, 14(4), 435–459. doi:10.1080/09541440143000159.

Murray, W., & Forster, K. I. (2004). Serial mechanisms in lexical access: the rank hypothesis. *Psychological Review*, *111*(3), 721–56. doi:10.1037/0033-295X.111.3.721

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: new evidence from the English Lexicon Project. *Psychonomic Bulletin Review*, *13*(1), 45–52.

Newman, R. S., & German, D. J. (2005). Life span effects of lexical factors on oral naming. *Language and Speech*, *48*(Pt 2), 123–56.

Norman, S., Kemper, S., & Kynette, D. (1992). Adults' reading comprehension: Effects of syntactic complexity and working memory. *Journal of Gerontology*, *47*(4), 258–265.

Norris, D. (2006). The Bayesian reader: explaining word recognition as an optimal Bayesian decision process. *Psychological review, 113*(2), 327.

O'Malley, S., & Besner, D. (2013). Reading aloud: Does previous trial history modulate the joint effects of stimulus quality and word frequency? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *39*(4), 1321–5. doi:10.1037/a0031673

Ossher, L., Flegal, K. E., & Lustig, C. (2012). Everyday memory errors in older adults. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, 37–41. doi:10.1080/13825585.2012.690365

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology*, *61*(2), 106-151.

Plaut, D. C., McClelland, J. L., Seidenberg, M., & Patterson, K. E. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review*, *103*(1), 56–115.

Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the lexical-decision task. *Psychology and Aging*, *19*(2), 278–89. doi:10.1037/0882-7974.19.2.278

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. (2012). *The psychology of reading*. Psychology Press.

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, *21*(3), 448–65. doi:10.1037/0882-7974.21.3.448

Rayner, K., Yang, J., Castelhano, M. S., & Liversedge, S. (2011). Eye movements of older and younger readers when reading disappearing text. *Psychology and Aging*, *26*(1), 214–23. doi:10.1037/a0021279

Salthouse, T. (1994). The nature of the influence of speed on adult age differences in cognition. *Developmental Psychology, 30*(2), 240–259.

Salthouse, T. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*(3), 403–28.

Salthouse, T. (2000). Methodological assumptions in cognitive aging research. *The Handbook of Aging and Cognition*.

Salthouse, T. (2004). What and When of Cognitive Aging. *Current Directions in Psychological Science, 13*(4), 140–144.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*(5), 763.

Samuels, S., LaBerge, D., & Bremer, C. (1978). Units of word recognition: Evidence for developmental changes. *Journal of Verbal Learning and Verbal Behavior*, *17*(6), 715–720.

Seidenberg, M. S. & McClelland, J. L. (1989) A distributed, developmental model of word recognition and naming. *Psychological Review,* 96, 523–68.

Schieber, F. (2006). Vision and Aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the Psychology of Aging* (6th ed.).

Schilling, H. H., Rayner, K., & Chumbley, J. (1998). Comparing naming, lexical decision, and eye fixation times: word frequency effects and individual differences. *Memory & Cognition*, *26*(6), 1270–81.

Schock, J., Cortese, M. J., & Khanna, M. M. (2012). Imageability estimates for 3,000 disyllabic words. *Behavior Research Methods*, *44*(2), 374-379.

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

Sommers, M. S., & Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in young and older adults: the interaction of lexical competition and semantic context. *Psychology and Aging*, *14*(3), 458–72.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science,* 411-416.

Spieler, D. H., & Balota, D. A. (2000). Factors influencing word naming in younger and older adults. *Psychology and Aging*, *15*(2), 225–231. doi:10.1037//0882-7974.15.2.225

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, *16*, 32-71.

Starns, J. J., & Ratcliff, R. (2010). The effects of aging on the speed–accuracy compromise: Boundary optimality in the diffusion model. *Psychology and aging*, *25*(2), 377.

Tainturier, M., Tremblay, M., & Lecours, A. R. (1989). Aging and the word frequency effect: A lexical decision investigation. *Neuropsychologia*, *27*, 1197–1203.

Thomas, J. C., Fozard, J. L., & Waugh, N. C. (1977). Age-related differences in naming latency. *The American Journal of Psychology*, *90*(3), 499–509.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology. General*, *124*(2), 107–36.

Tun, P., McCoy, S. L., & Wingfield, A. (2009). Aging, hearing acuity, and the attentional costs of effortful listening. *Psychology and Aging*, *24*(3), 761–6. doi:10.1037/a0014802

Unsworth, N., Redick, T.S., McMillan, B.D., Hambrick, D.Z., Kane, M.J., & Engle, R.W. (2015) Is playing video games related to cognitive abilities? *Psychological Science.* doi: 10.1177/0956797615570367.

Van der Linden, M., Hupet, M., Feyereisen, P., Schelstraete, M.-A., Bestgen, Y., Bruyer, R., … Seron, X. (1999). Cognitive mediators of age-related differences in language comprehension and verbal memory performance. *Aging, Neuropsychology, and Cognition*, *6*(1), 32–55.

Verhaeghen, P. (2003). Aging and vocabulary score: A meta-analysis. *Psychology and Aging*, *18*(2), 332–339. doi:10.1037/0882-7974.18.2.332

Verhaeghen, P., & Salthouse, T. (1997). Meta-analyses of age-cognition relations in adulthood: estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin*, *122*(3), 231–49.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*, 1191–207. doi:10.3758/s13428-012-0314-x

Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review*, *103*(4), 761–72.

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *The Quarterly Journal of Experimental Psychology: Section A*, *50*(2), 439-456.

Whaley, C. P. (1978). Word—nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, *17*(2), 143–154. doi:10.1016/S0022-5371(78)90110-X

Whiting, W. L., Madden, D. J., Langley, L. K., Denny, L. L., Turkington, T. G., Provenzale, J. M., … Coleman, R. E. (2003). Lexical and sublexical components of age-related changes in neural activation during visual word identification. *Journal of Cognitive Neuroscience*, *15*(3), 475–87. doi:10.1162/089892903321593171

Whitney, C., & Cornelissen, P. (2008). SERIOL reading. Language and Cognitive Processes, 23, 143–164. doi:10.1080/01690960701579771

Wingfield, A., Tun, P., & McCoy, S. L. (2005). Hearing loss in older adulthood: What it is and how it interacts with cognitive performance. *Current Directions in Psychological Science, 14*(3), 144–148. doi:10.1111/j.0963-7214.2005.00356.x

Wingfield, A., Waters, G. S., & Tun, P. (1998). Does working memory work in language comprehension?: Evidence from behavioral neuroscience. In N. Raz (Ed.), *The Other Side of the Error Term - Aging and Development as Model Systems in Cognitive Neuroscience* (Vol. 125, pp. 319–393). Amsterdam: Elsevier. doi:10.1016/S0166-4115(98)80009-5

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, *60*(4), 502–529. doi:10.1016/j.jml.2009.02.001

Yap, M., & Balota, D. (2015). Visual Word Recognition. In *The Oxford Handbook of Reading* (pp. 1–22). Oxford University Press. doi:10.1093/oxfordhb/9780199324576.013.4

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: insights from the English Lexicon Project. *Journal of Experimental Psychology. Human Perception and Performance*, *38*(1), 53–79. doi:10.1037/a0024177

Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., & Huff, M. J. (2012). An abundance of riches: cross-task comparisons of semantic richness effects in visual word recognition. *Frontiers in Human Neuroscience*, *6*(April), 72. doi:10.3389/fnhum.2012.00072

Yap, M. J., Tse, C.-S., & Balota, D. A. (2009). Individual differences in the joint effects of semantic priming and word frequency: The role of lexical integrity. *Journal of Memory and Language*, *61*(3), 303. doi:10.1016/j.jml.2009.07.001

Yarkoni, T., Balota, D. A., & Yap, M. J. (2008). Moving beyond Coltheart's N: a new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*(5), 971–9. doi:10.3758/PBR.15.5.971

Zeno, S.M., Ivens, S.H., Millard, R.T., & Duvvuri, R. (1995). The educator's word frequency guide. Brewster, NY: Touchstone Applied Science.

Zevin, J. D., & Seidenberg, M. (2002). Age of acquisition effects in reading aloud: Tests of cumulative frequency and frequency trajectory. *Memory & Cognition*, *47*, 1–29.

Ziegler, J. C., Petrova, A., & Ferrand, L. (2008). Feedback consistency effects in visual and auditory word recognition: where do we stand after more than a decade? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*(3), 643–61. doi:10.1037/0278-7393.34.3.643

Appendices

Appendix A: Word Stimuli

acrobat
actor
addict
admiral
admirer
adult
affinity
aficionado
agenda
aid
aisle
albatross
alcohol
alderman
alfalfa
algorithm
alligator
almanac
alphabet
ambition
amplifier
anaconda
analyst
anecdote
anemone
angle
animal
ant
antacid
anteater
antelope
ape
aphid
apparatus
apprentice
apricot
arcade
arch
architect
argument
art
artichoke
artisan
artist
asphalt
assassin
assistant
asylum
athlete
attacker
attorney
auditor
aunt
author
avalanche
avenue
aviator
avocado
award
awning
axe
baboon
baby
bachelor
badger
bagpipe
baker
balcony
ball
ballerina
bamboo
barber
barnacle
baron
barrack
barracuda
basket
bath
battle
beacon
bear
beast
beatnik
beaver
bed
bee
beer
beetle
bib
bigot
billiard
bingo
bird
bishop
bison
block
bluebird
boa
boar
boatman
body
bodyguard
bomb
bonus
boomerang
booze
bottle
bourbon
bow
boy
brain
brake
brass
brasserie
brat
breach
bride
brisket
broccoli
bronco
brooch
brother
brunette
brute
buccaneer
buffalo
buffoon
bull
bully
bunk
bunny
burglar
bush
butcher
butler
button
cabbage
cable
caddie
cadet
cage
calf
caliber
caller
calm
camel
campaign
canary
candidate
cane
canister
cannibal
cannon
canoe
canopy
canteen
capsule
captain
caption
carafe
cardigan
cardinal
caretaker
cargo
caribou
caricature
carp
carpenter
cart
carton
casket
cassette
cat
cataclysm
caterpillar
catfish
cauliflower
cave
ceiling
celebrity
celery
cell
cellar
cellist
cement
cent
chairman
challenge
chamber
chameleon
champion
chap
chapel
chaplain
charlatan
cheetah
chef
chess
chick
chicken
chief
chieftain
child
chimpanzee
chinchilla
chipmunk
chocolate
chord
chrysanthemum
church
chutney
cicada
cider
circle
citizen
clang
cleric
clerk
client
cloak
cloud
clover
clown
clue
cobalt
cobra
cockatoo
cod
code
codfish
codger
collar
colleague
collie
colonel
columnist
commander
comment
commerce
commuter
compact
compartment
compass
component
composer
comrade
concourse
concubine
condition
consonant
consumer

contract	curve	doll	entree	flower	glimmer	halibut
contraption	customer	dolphin	entrepreneur	fluff	gnat	hammer
copper	cyclist	donkey	epiphany	fluid	goal	hamster
copy	czar	door	errand	foal	goat	harbor
cord	dachshund	doorkeeper	estate	foe	gold	hare
coroner	dad	dove	ewe	fog	golf	harlot
cosmetic	daffodil	dowel	example	folk	golfer	harpoon
cottage	dalmatian	dozen	exile	fondue	goon	hassle
cougar	dame	draft	expense	font	goose	havoc
countess	damsel	drama	expert	fork	gopher	hawk
coupon	dandelion	drape	extravaganza	foundry	gorilla	hawker
court	dart	dress	facade	fowl	goulash	hazard
cousin	dash	drift	fact	fox	grade	hearse
cow	daughter	drizzle	falcon	frame	grandmother	heifer
coyote	deacon	drug	famine	freight	grandson	height
cradle	debit	drummer	fang	friar	granite	heir
craftsman	debris	duck	farmer	frigate	granule	hemlock
crate	debutante	duct	fate	fright	grape	hen
creature	deceit	duffel	father	frog	grate	hermit
creek	deck	duke	faucet	fudge	greed	heroine
crest	deed	dunce	fauna	fuel	grief	herring
cricket	deer	dusk	fawn	fund	gristle	hiatus
crime	dent	dwarf	fellow	furniture	grocer	hibiscus
critic	depth	eagle	felon	fuss	grouch	hippie
crocodile	diamond	earl	female	gadget	grouse	hobo
crony	dingo	easel	fern	gal	guard	hockey
crook	dinner	edge	ferret	gangster	guerilla	hog
crow	dirt	editor	fiancee	gap	guest	hoodlum
crunch	disaster	eel	fiddle	garage	guillotine	hooligan
crusader	discipline	elephant	fiend	gate	gull	hoop
crypt	disdain	elk	fillet	gaze	gumption	horn
cub	dish	elm	filter	gazelle	gunman	hornet
cubicle	ditch	emblem	finale	gecko	guppy	horse
cuff	diva	emperor	fire	geezer	guru	hostage
culprit	divorce	emphysema	fish	geisha	gutter	hour
cup	docket	emu	flamingo	gentleman	guy	hue
curry	doctor	endeavor	flap	geranium	gymnasium	human
curse	document	enemy	flea	ghetto	gymnast	humility
curtain	doe	engineer	flight	gig	hacksaw	hunter
curtsy	dog	entrance	florist	girl	haiku	husband

hut
hyacinth
hydrangea
hyena
hyphen
icicle
idea
idiot
iguana
impostor
index
inmate
innuendo
insect
interval
intestine
intruder
inventory
invoice
irony
issue
italic
jackal
jacket
jaguar
jail
janitor
jar
jaunt
javelin
jazz
jeans
jet
jewel
job
jockey
judge
juice
juror
kale
kangaroo
keg
kelp
kerchief
kerosene
kid
kidnapper
kidney
kiln
king
kiosk
kitchen
kite
kitten
kiwi
knight
koala
kudos
kumquat
label
labor
laboratory
lacquer
lad
ladder
ladle
lady
lamb
lanyard
lark
larynx
lass
latch
lattice
lava
lawyer
leash
leather
lecturer
ledge
ledger
lemur
length
leopard
leper
letter
library
lice
license
lieutenant
limb
limbo
limousine
line
linen
lingo
linoleum
lint
lion
lioness
lipstick
liquid
liquor
list
literature
lizard
llama
load
loaf
lobster
location
locust
lodge
logger
loophole
lotion
lottery
lounge
lung
luxury
machete
machine
machinist
mackerel
madam
maggot
magnet
magnolia
magpie
maid
mailbox
mallard
mammal
mammoth
man
manatee
maniac
manor
manual
marathon
mare
marigold
market
marsupial
mart
marvel
mascara
master
mastiff
matador
match
matron
mausoleum
maverick
mayonnaise
mayor
medic
medicine
mentor
merchant
mermaid
mesh
message
messenger
metal
meteor
methane
microwave
midwife
migrant
mile
milestone
mill
mime
minister
minx
mirror
miscreant
miser
mister
mistress
mite
mitten
moisture
molasses
mom
mongoose
monk
monkey
monsoon
month
moon
moose
mortal
mosquito
moth
mother
motorist
mouse
mud
mule
mushroom
mutt
mystery
nanny
narcotic
needle
neighbor
nephew
newt
nickel
niece
noise
nomad
noon
note
notice
nozzle
nudge
nudist
nun
nurse
nymph
oaf
oak
oatmeal
obsidian
occasion
octave
octopus
odor
officer
onion
onslaught
opossum
opponent
optics
option
orchid
origin
oriole
orphan
osprey
ostrich
otter
outburst
oven
overtime
owl
ox
oyster
paddock
painter
pamphlet
panda
parachute
paragraph
parakeet
parent
parole
parrot
parsley
participle
partner
partridge
passenger
passport
paste
pastor
patron
pauper
pavilion
payday
peacock
peasant
pedestal
pedestrian
peek
peg
pelican

penance
pendant
penguin
pennant
pepper
person
personnel
pest
pet
petal
petunia
pharaoh
pharmacy
pheasant
phrase
picket
pig
pigeon
pilgrim
pillar
pillow
pilot
pinball
pioneer
piranha
pistol
pitch
pizza
place
plagiarism
plane
plankton
plastic
plate
plateau
platter
platypus
playboy
plumber
plywood
pocket
poem
poet
poinsettia
police
polo
pony
poodle
pool
porcupine
porpoise
porridge
pothole
powder
prawn
precision
premise
prestige
prey
priest
primate
prince
princess
principal
prisoner
promenade
propane
prostitute
prowl
pub
puff
puffin
puma
pumpkin
pup
pupil
python
quail
quarry
queen
quest
quill
rabbi
rabbit
raccoon
radish
raffle
ranger
ransom
rat
raven
realm
realtor
rebel
referee
refugee
reign
rent
reptile
respect
restaurant
reward
rhinoceros
rhubarb
ride
rider
ridge
rift
ring
river
roach
road
robot
rocket
rodent
rookie
roster
routine
rubber
rumble
rut
saboteur
sack
sailor
salamander
salary
salmon
samaritan
samurai
sardine
satellite
saucepan
sausage
savant
saxophone
scarf
scat
scenario
scientist
scone
scoop
scorn
scorpion
scrapbook
screech
script
seahorse
seance
secrecy
secretary
senator
serenade
sergeant
serpent
servant
shade
shaft
shark
shawl
sheep
shell
shellac
shepherd
sheriff
ship
shoelace
shopper
shotgun
shrapnel
shred
shredder
shrew
shriek
shrimp
shrine
shrub
shuffle
shutter
sight
silicon
silo
simile
singer
sister
skirt
skunk
sleuth
sliver
slob
sloth
sludge
slug
smog
smudge
snake
sniper
snitch
snob
soap
soccer
solace
soldier
son
sonata
soot
sophomore
sorrow
spa
space
spaghetti
spaniel
spider
spinach
spinster
spiral
splash
spouse
spout
spray
spring
spy
squat
squaw
squid
squire
squirrel
stack
stag
stalagmite
stallion
staple
statue
stench
stepson
steward
stiletto
stilts
stirrup
stockade
store
stove
stranger
strap
stretcher
stride
strife
stucco
student
stunt
stupor
sturgeon
suit
suitor
sultan
summons
supply
suture
sweat
swine
swirl
sycamore
syllable
symbol
symphony
syringe
system
tab
tactic
tadpole
tailor
tale
tambourine
taper
tapestry
tarantula
tarot
teacher
teal
technique
tee

teen	tinge	treaty	turret	vegan	watchdog	witch
teenager	tinkle	tree	turtle	velvet	wave	witness
television	toast	trend	tutor	vermin	way	wolf
temple	toll	trilogy	tuxedo	vermouth	weasel	woman
terracotta	tomato	trinket	tweezer	vertigo	wedge	wombat
terrier	tongue	tripod	twist	vest	welt	worker
test	tortoise	triumph	tycoon	villain	whale	worm
thespian	tote	trombone	umpire	viper	whimper	wrench
thief	toucan	trouble	uncle	visitor	whistle	yacht
thigh	towel	trout	urchin	vulture	widow	yak
throne	tower	trowel	urge	waiter	wife	zebra
thug	trail	trunk	vagabond	wall	wig	zest
tidbit	trait	tug	valor	wallaby	wildebeest	zone
tiger	tram	tuna	varnish	walrus	willow	
tile	trash	turban	vase	wand	wink	
tin	traveller	turkey	vat	wasp	winter	

Appendix B: Nonword Stimuli

absitar	boam	coamin	eapal	fustery	jucur	mied
aggict	boddle	coclo	ebyl	gake	juilt	milestant
agolt	boopon	coft	eledrite	garepler	jut	milvife
ahrioner	borrant	cogrish	ence	gilmer	kenchen	miprand
aif	borry	combiss	enviteur	gisk	ker	mistopove
alcyment	brade	commetion	esdense	gisman	kershoof	moby
alfter	brage	compesher	esk	goddusium	kint	mocket
alipol	breek	compit	estine	goir	kipbipper	moel
alks	breeture	cond	etcition	gouse	lale	moep
allailer	bretter	congart	etinar	grackroiler	larl	mon
ansal	brip	consolite	etolyst	gumph	lattave	mongoan
assissite	brocken	constave	eubre	guoys	lawed	mormel
asynoy	burswor	corcoter	exaybra	guvi	layther	moshfure
atcurer	calbish	corraser	fanen	gysdast	leggarer	mowper
atserpan	candibent	cortnet	fap	hanbir	lemacs	muccer
attossin	cang	cothanee	fapide	hawd	lims	muntry
awplor	cank	cotmonic	faration	hesa	lind	mupit
baff	cannoun	cottide	fiel	het	lirebision	murvant
baglope	canranter	crilt	fiflory	hobelable	lomost	myrisse
balfer	capou	crortsman	fimpo	hunem	lonce	nedge
bample	carf	crub	fipprick	hunen	loroan	nelps
bancer	catiri	cugtrit	fiqoed	huon	loroism	nen
bapet	catofy	cunge	fireer	hustand	lourse	newe
battet	cend	cyscast	flape	icety	lyete	noits
beash	cetihan	dack	flie	iley	machent	nolive
beathok	cettower	dag	flooker	imed	mailrax	nulits
beb	chedney	delte	fluesird	imiatul	mambyth	nuodess
beddalo	chex	dend	flug	imitra	manent	nyard
bederatums	chibod	dimed	fof	ingell	maraflote	oatreel
beebon	chinfess	diseine	foft	inment	maramil	occovos
bettan	chinret	dollow	foucher	instoler	masp	ocyban
bibin	chird	dorey	fouther	isbant	mastion	onsce
bild	choof	dort	fow	itceror	materee	ool
billist	chupper	dosh	frief	jayurg	mautoceow	orm
bimed	clate	dotplin	fromber	jendel	mecket	ovein
bipoceom	clern	doudle	frusoler	jeninar	mercett	overmegm
bistom	clide	driffle	fudger	jol	mester	owelern
blandbon	cload	dule	fuke	jondle	micket	paddand
blug	clow	dybron	furgle	juckem	middor	pagor

pagsay
palode
palstret
pammel
panbon
panghy
pango
paraclern
paror
partbort
paspora
pawritist
peb
pector
pemmenger
pentol
pheckle
phome
pimed
pinlave
pirka
pirter
plagpoy
plenmacy
plimmer
pobbit
ponger
potnile
proost
pugma
punnant
puticine
rad
ramad
rareor
rasata
redite
reimond
restaprite
rewubia
rexid
rippian
runkey
rupper
russer
ruttery
sananar
saruet
saturist
schipnel
scoon
selicid
senuac
sergeesh
settable
sheese
sheilall
shess
shilgun
shollion
shremper
shrile
shuripe
shuzzle
sibscony
sieror
slayon
sleed
slout
sotsier
sporer
sposist
spuric
spustic
squag
stadaller
staw
steloby
stepand
storon
strinker
stryncher
stubant
stunoner
sudsage
sumbon
sunsey
swaper
tagy
tanchter
tarcoilin
tarmy
tassier
tauler
tawyon
terlan
toaracer
trasel
tupple
twealant
tymocatogy
uan
ucoa
uctall
ulor
ultion
umpond
uniporne
urswan
uscicer
varcunic
vegar
vetcine
wamdus
warpen
wartet
weebel
wembing
wesan
whill
wicack
wisinar
witop
woir
zenacs
zetten
ziope

Appendix C. Demographic Questionnaire

1. Please report your age (in years):___________
2. Please report your gender:______________
3. Please indicate the number of years of schooling that you have completed. _____
(12 = finished high school, add or subtract years for more/less education)
4. Please report your field of study/school major or occupation: ___________________

5. Please indicate what time of day you feel most alert:
_____Morning
_____Afternoon
_____Evening
_____No differences

6. Please place a check beside one or more of the following racial categories that apply to you:
_____ American Indian / Alaskan Native
_____Asian
_____Native Hawaiian or Other Pacific Islander
_____Black / African American
_____White / Caucasian
_____Prefer Not to Respond

7. Do you consider yourself to be Hispanic or Latino?
____Yes
____No
____Prefer Not to Respond

8. Is English your first language?
___Yes
___No → Please indicate your first language________________

9. Is there anything we should know about that might affect your performance during the testing session today (e.g. lack of sleep, feeling ill, etc.)?

10. How would you describe your socioeconomic status relative to society?
(a) Significantly above average
(b) Above average
(c) Average
(d) Below average

(e) Significantly below average

11. What is your approximate income bracket? If currently retired, select income during past employment. If dependent, select household income bracket.
(a) $0- $19,999
(b) $20,000- $49,999
(c) $50,000- $79,999
(d) $80,000- $109,999
(e) $110,000- $139,999
(f) $140,000- $169,999
(g) $170,000- $199,999
(h) $200,000+

12. How often do you use a computer? (Please report hours per day and days per week)

____

13. How comfortable and proficient do you feel with using a computer for normal daily use (e.g., not advanced/programming functions), from 0 (very uncomfortable, not at all proficient) to 10 (totally at ease, completely proficient)? _____

14. How often do you read? (Please report hours per day and days per week)

____

15. Which format do you read? (Check all that apply)
____ In print (hard copy)
____ Electronically (on a computer)
____ Other: ________________

16. Do you feel that your reading habits have changed at all over your lifetime? If yes, how so?